

# ADTSC Science Highlights 2013

Associate Directorate for Theory, Simulation, and Computation

## *Preface*

The Theory, Simulation, and Computation Directorate encompasses a very broad array of technical disciplines. However, our goal is to integrate across these disciplines and work with our experimental colleagues to develop a truly predictive understanding of the complex problems the country faces in national security. Combining validated theory, algorithmic developments, and high-performance computing, we can raise the notion of the scientific method to Information, Science, and Technology for Prediction.

This annual publication presents representative examples of the interdisciplinary research that we undertake in ADTSC, partnering with our colleagues from across the Laboratory, and indeed from across the globe, toward “Integrating Information, Science, and Technology for Prediction.”

Paul J. Dotson

*Associate Director (acting), Theory, Simulation, and Computation Directorate*

Paul J. Dotson  
*Associate Director (acting)*

Audrey L. Archuleta  
*ADTSC Chief of Staff*

Stephen R. Lee  
*CCS Division Leader*

Gary Grider  
*HPC Division Leader (acting)*

Antonio Redondo  
*T Division Leader*

## **Publications Team**

Sharon Mikkelsen  
*Design and Coordination*

Kathy Pallis  
*Writing/Editing*



# Contents

## Applied Mathematics and Fluid Dynamics

<b>High-Fidelity Simulations of Turbulent Flow In Nuclear Reactors.....</b>	<b>4</b>
Mark A. Christon, Jozef Bakosi, Robert B. Lowrie, Lori Pritchett-Sheats, CCS-2; Robert R. Nourgaliev, Idaho National Laboratory	
<b>Modeling Fluid Flow in Domains Containing Moving Interfaces .....</b>	<b>6</b>
David B. Carrington, T-3; Dominic A. Muñoz, Juan C. Heinrich, University of New Mexico	
<b>Adaptive Finite Elements with local ALE for Modeling Turbulent Reactive Flow in Engines with Injection.....</b>	<b>8</b>
David B. Carrington, T-3; Xiuling Wang, Purdue University; Juan C. Heinrich, University of New Mexico; Darrell W. Pepper, University of Nevada	
<b>M-Adaptation for Acoustic Wave Equation in 3D.....</b>	<b>10</b>
Vitaliy Gyrya, Konstantin Lipnikov, T-5	
<b>Late-Time Quadratic Growth in Single-Mode Rayleigh-Taylor Instability.....</b>	<b>12</b>
Daniel Livescu, Tie Wei, CCS-2	
<b>Direct Numerical Simulations of Rayleigh-Taylor Instability with Gravity Reversal.....</b>	<b>14</b>
Daniel Livescu, Tie Wei, CCS-2	
<b>Turbulence Model Validation for Complex Mixing Scenarios.....</b>	<b>16</b>
Bertrand Rollin, CCS-2; Nicholas A. Denissen, Jon M. Reisner, Malcolm J. Andrews, XCP-4; Jimmy Fung, XCP-1	
<b>Towards a Next-Generation Reconnection-ALE Hydrocode .....</b>	<b>18</b>
Shiv K. Sambasivan, CCS-2; Mikhail J. Shashkov, XCP-4	
<b>Minimal Stencil Finite Volume Scheme with the Discrete Maximum Principle .....</b>	<b>20</b>
Konstantin Lipnikov, Daniil Svyatskiy, T-5; Yuri Vassilevki, INM Russian Academy of Sciences, Moscow, Russia	

## Astrophysics and Cosmology

<b>Studying Supernova Progenitors with Observations and Simulations .....</b>	<b>24</b>
Lucille H. Frey, HPC-3, University of New Mexico; Wesley Even, XTD-6; Amanda J. Bayless, Southwest Research Institute; Chris L. Fryer, CCS-2, University of Arizona, University of New Mexico; Peter W.A. Roming, Southwest Research Institute, Pennsylvania State University; Tyler Pritchard, Pennsylvania State University	
<b>Supernova Remnants: Nature's Turbulence Experiment.....</b>	<b>26</b>
Chris L. Fryer, CCS-2; Carola Ellinger, University of Texas; Andreas Zoglauer, University of California	
<b>The First Planets .....</b>	<b>28</b>
Jarrett L. Johnson, Hui Li, T-2	



<b>A Strong Case for Supermassive Stars in the Early Universe.....</b>	<b>30</b>
Jarrett L. Johnson, T-2; Daniel J. Whalen, Carnegie Mellon University; Hui Li, T-2; Claudio Dalla Vecchia, Bhaskar Agarwal, Sadegh Khochfar, Max Planck Institute for Extraterrestrial Physics; Daniel E. Holz, University of Chicago	
<b>The Hashed Oct-Tree N-Body Algorithm at a Petaflop .....</b>	<b>32</b>
Michael S. Warren, T-2; Ben Bergen, CCS-7	
 <b>Climate, Atmospheric, and Earth Systems Modeling</b>	
<b>Scale-Aware Sub-Grid Eddy Parametrizations for Ocean Modeling.....</b>	<b>36</b>
Qingshan Chen, Todd Ringler, T-3	
<b>Uncertainty Quantification and Ocean Model Analysis .....</b>	<b>38</b>
James R. Gattiker, Sham Bhat, CCS-6; Matthew W. Hecht, CCS-2	
<b>Level-Ice Melt Ponds in the LANL Sea Ice Model, CICE .....</b>	<b>40</b>
Elizabeth C. Hunke, T-3; David Hebert, Naval Research Laboratory, Stennis Space Center; Olivier Lecomte, Université Catholique de Louvain, Belgium	
<b>Glaciers' Response in a Warming Climate .....</b>	<b>42</b>
Sebastian H. Mernild, CCS-2	
<b>The Arctic Terrestrial Simulator: Modeling Permafrost Degradation in a Warming Arctic.....</b>	<b>44</b>
Ethan Coon, Rao Garimella, Gianmarco Manzini, J. David Moulton, Daniil Svyatskiy, T-5; Markus Berndt, CCS-2; Scott Painter, EES-16	
<b>On a Turbulent Energy Transfer Scale in Ocean Turbulence .....</b>	<b>46</b>
Balu T. Nadiga, CCS-2; David N. Straub, McGill University	
<b>Statistical Mechanics of the Geometric Control of Flow Topology in 2D Turbulence.....</b>	<b>48</b>
Balu.T. Nadiga, CCS-2; Peter N. Loxley, T-5	
<b>A Three-Dimensional Eddy Census of a High-resolution Global Ocean Simulation.....</b>	<b>50</b>
Mark R. Petersen, CCS-2; Sean J. Williams, CCS-3; Matthew W. Hecht, CCS-2; Mathew E. Maltrud, T-3; Bernd Hamann, University of California	
<b>A Multi-Resolution Approach to Global Ocean Modeling .....</b>	<b>52</b>
Mark R. Petersen, CCS-2; Todd Ringler, Doug Jacobsen, Mathew E. Maltrud, Philip W. Jones, T-3	
<b>Two Modes of Sea-Ice Gravity Drainage.....</b>	<b>54</b>
Adrian K. Turner, Elizabeth C. Hunke, T-3; Cecilia M. Bitz, University of Washington	
<b>Modal Variability in the Bellingshausen Basin .....</b>	<b>56</b>
Wilbert Weijer, CCS-2	
<b>Unified, Unstructured Grids for Climate Modeling.....</b>	<b>58</b>
Geoffrey A. Womeldorff, T-3; Janet Peterson, Max Gunzburger, Florida State University; Todd Ringler, T-3	



## Information Science and Technology

<b>Strategic Data Collection for Optimal Use of Resources Based on Multiple Criteria .....</b>	<b>62</b>
Christine Anderson-Cook, CCS-6	
<b>Non-Hermitian Approach for Modeling of Noise-Assisted Quantum Electron Transfer in Photosynthetic Complexes .....</b>	<b>64</b>
Gennady P. Berman, T-4; Alan R. Bishop, PADSTE; Alexander I. Nesterov, Universidad de Guadalajara	
<b>Exploring the Possibility of Quantitative Roles for Process Monitoring in Nuclear Safeguards .....</b>	<b>66</b>
Thomas L. Burr, Michael S. Hamada, Mikhail Skurikhin, Brian Weaver, CCS-6	
<b>An Innovative Parallel Cloud Storage System using OpenStack's Swift Object Store and Transformative Parallel I/O Approach .....</b>	<b>68</b>
Hsing-Bung Chen, Benjamin McClelland, HPC-5; David Sherrill, HPC-3; Alfred Torrez, HPC-1; Parks Fields, HPC-5; Pamela Smith, HPC-3	
<b>iBridge: Improving Unaligned Parallel File Access with Solid-State Drives .....</b>	<b>70</b>
Kei Davis, CCS-7; Xuechen Zhang, Georgia Institute of Technology; Ke Liu, Song Jiang, Wayne State University	
<b>Improving Disk Performance: A Prefetching Scheme Exploiting Data Layout and Access History .....</b>	<b>72</b>
Kei Davis, CCS-7; Xiaoning Ding, Xiaodong Zhang, The Ohio State University; Song Jiang, Wayne State University	
<b>Making Memory Swapping Practicable: Synergistic Coupling of SSD and Hard Disk for QoS-Aware Virtual Memory .....</b>	<b>74</b>
Kei Davis, CCS-7; Song Jiang, Wayne State University; Xuechen Zhang, Georgia Institute of Technology; Ke Liu, Wayne State University	
<b>Reliability Models for Double Chipkill Detect/Correct Memory Systems .....</b>	<b>76</b>
Nathan DeBardeleben, Sean Blanchard, HPC-5; Rakesh Kumar, Stevenson Jian, University of Illinois; Vilas Sridharan, Advanced Micro Drives	
<b>Network Traffic Generator for Cyber Security Testbeds .....</b>	<b>78</b>
Hristo Djidjev, CCS-3; Lyudmil Aleksandrov, Bulgarian Academy of Sciences	
<b>PENNANT: A Research Tool for Unstructured Mesh Physics on Advanced Architectures .....</b>	<b>80</b>
Charles R. Ferenbaugh, HPC-1	
<b>Parallel Log Structured File System (PLFS) .....</b>	<b>82</b>
Gary Grider, HPC-DO; John Bent, EMC Corporation; Chuck Cranor, Carnegie Mellon University; Jun He, New Mexico Consortium; Aaron Torres, HPC-3; Meghan McClelland, Brett Kettering, HPC-5	
<b>Yellow to Turquoise Integration (YeTI) Project .....</b>	<b>84</b>
Brett Kettering, Benjamin McClelland, HPC-5; Kyle Lamb, HPC-3; Alex Malin, HPC-DO	
<b>Epidemic Modeling and Facemask Usage.....</b>	<b>86</b>
Susan M. Mniszewski, CCS-3; Sara Y. Del Valle, Reid Priedhorsky, D-4; James M. Hyman, Kyle S. Hickman, Tulane University	
<b>Co-Design for Molecular Dynamics: An Exascale Proxy Application .....</b>	<b>88</b>
Jamaludin Mohd-Yusof, Sriram Swaminarayan, CCS-7; Timothy C. Germann, T-1	

<b>Accelerating Graph Algorithms Using Graphics Processors: Shortest Paths for Planar Graphs .....</b>	<b>90</b>
Hristo Djidjev, Sunil Thulasidasan, CCS-3; Guillaume Chapuis, Rumen Andonov, University of Rennes, France	
<b>Network Uncertainty on Electric Power Grids .....</b>	<b>92</b>
Russell Bent, D-4; Earl Lawrence, Scott Vander Wiel, CCS-6	
<b>A Computational Social Model of the Education System.....</b>	<b>94</b>
Stephan Eidenbenz, Patrick Kelly, CCS-3; Benjamin Sims, Joanne Wendelberger, CCS-6	

## Atomic, Nuclear, and High-Energy Physics

<b>Radiation Damping for Speeding Up NMR Applications.....</b>	<b>98</b>
Gennady P. Berman, T-4; Michelle A. Espy, Petr L. Volegov, P-21; Vyacheslav N. Gorshkov, National Technical University of Ukraine, Vladimir I. Tsifrinovich, Polytechnic Institute of New York University	
<b>Flavored Baryogenesis .....</b>	<b>100</b>
Vincenzo Cirigliano, Christopher Lee, T-2	
<b>Light Element Opacities of Astrophysical Interest from ATOMIC .....</b>	<b>102</b>
James P. Colgan, David Kilcrease, Norman H. Magee, Jr., Gregory Armstrong, Joseph Abdallah, Jr., Manolo Sherrill, T-1; Christopher J. Fontes, Honglin Zhang, Peter Hakel, XCP-5	
<b>Using Neutron Star Observations to Constrain Nuclear Physics .....</b>	<b>104</b>
Joseph A. Carlson, Stefano Gandolfi, T-2	
<b>Direct Numerical Tests of Kinetic Theory Collision Integrals with Molecular Dynamics Simulations of Stopping Power in Plasmas .....</b>	<b>106</b>
Paul E. Grabowski, Michael S. Murillo, CCS-2; Michael P. Surh, David F. Richards, A. Bruce Langdon, Frank R. Graziani, Lawrence Livermore National Laboratory	
<b>The New Finite-Range Droplet-Model Mass Table FRDM(2012) and Associated Tables of <math>\beta</math>-Decay Properties .....</b>	<b>108</b>
Peter Möller, T-2	

## Chemistry and Biology

<b>New Insights into Protein-DNA Binding from Simulations of DNA Breathing Dynamics .....</b>	<b>112</b>
Boian S. Alexandrov, T-1; Amy L. Bauer, XTD-4; Alan R. Bishop, PADSTE; William S. Hlavacek, Fangping Mu, T-6; Elizabeth Hong-Geller, B-10; Kristy Nowak-Lovato, D-3; Kim Ø. Rasmussen, T-5; Ludmil B. Alexandrov, Wellcome Trust Sanger Institute; Anny Usheva, Harvard Medical School	
<b>Epidemiological Analysis for Outbreak Control .....</b>	<b>114</b>
Nick Hengartner, CCS-3; Ruy M. Ribeiro, Paul W. Fenimore, T-6	
<b>Making the Dark Matter of Biology Visible .....</b>	<b>116</b>
Joel Berendzen, P-21; Mira Bussod, T-6; Judith Cohn, Nick Hengartner, CCS-3; Ben McMahon, T-6	

<b>Time-of-Infection Estimated by a Time-Continuous HIV-1 IgG Model .....</b>	<b>118</b>
Helena Skar, Thomas Leitner, T-6	
<b>Insights into Microscopic Diffusion Processes at a Solid/Fluid Interface under Supercritical Conditions: A Study of the Aqueous Calcite (10<math>\bar{1}</math>4) Surface .....</b>	<b>120</b>
Chun-Yaung Lu, Danny Perez, T-1; Donald D. Hickmott, TT-DO; Arthur F. Voter, T-1	
<b>Close Integration between Theory and Experiment: Simulations and Chemical Probing Experiments of Molecular Switches .....</b>	<b>122</b>
Karissa Sanbonmatsu, Scott Hennelly, T-6; Jose Onuchic, Ryan Hayes, Rice University	
<b>Computational Study of Electrochemical Windows of Room-Temperature Ionic Liquids .....</b>	<b>124</b>
Yong-Hui Tian, T-1; George S. Goff, Wolfgang H. Runde, C-IIAC; Enrique R. Batista, T-1	

## Materials Science

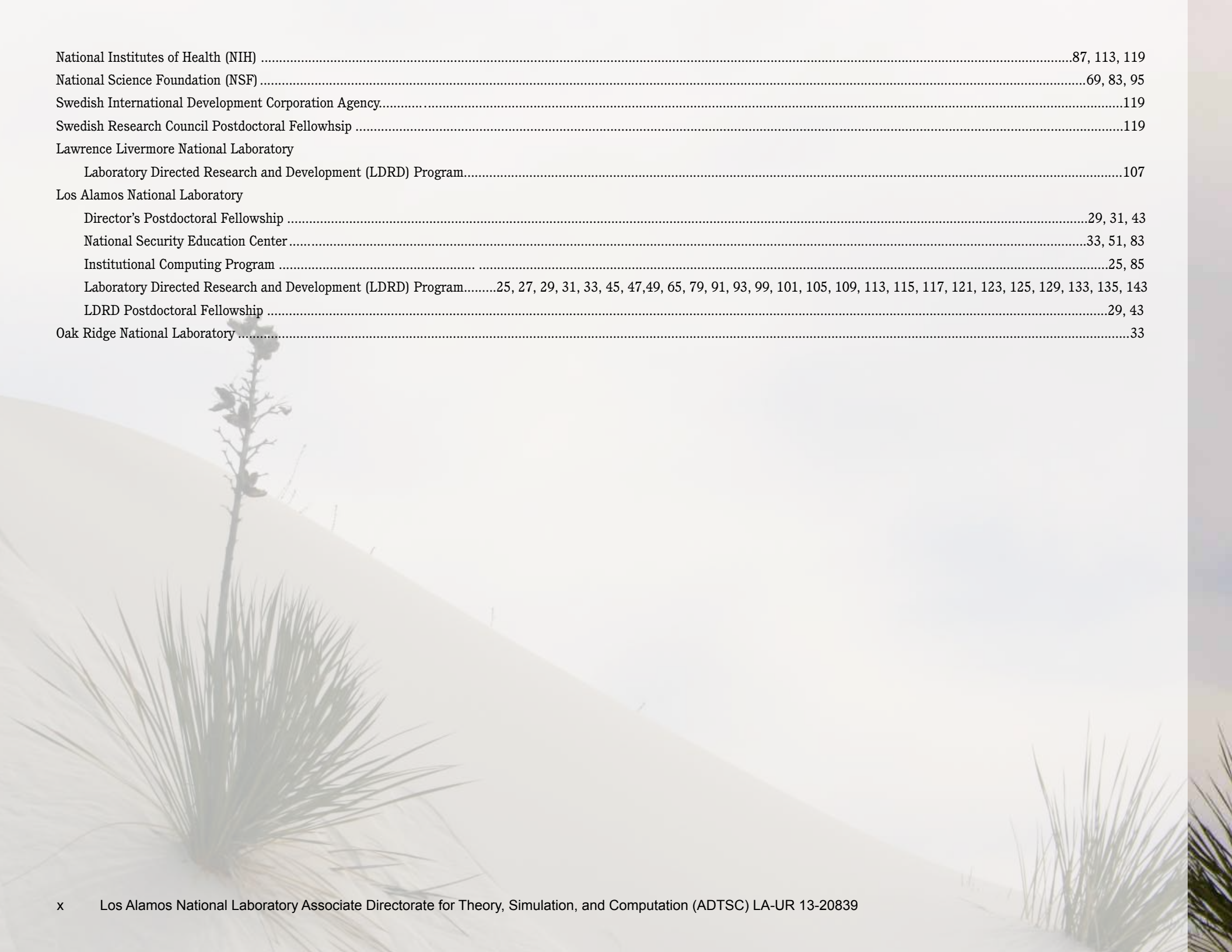
<b>Quantum-Based Molecular Dynamics on Graphics Processing Units .....</b>	<b>128</b>
Susan M. Mniszewski, CCS-3; Anders Niklasson, Ed Sanville, Marc J. Cawkwell, T-1	
<b>Thermally Induced Damage in Energetic Materials .....</b>	<b>130</b>
Bradford E. Clements, Axinte Ionita (T-1)	
<b>Modeling the Texture Evolution of Cu/Nb Layered Composites During Rolling.....</b>	<b>132</b>
Benjamin L. Hansen, T-3; John S. Carpenter, MST-6; Stephen D. Sintay, IAT-3; Curt A. Bronkhorst, T-3; Rodney J. McCabe, MST-6; Jason R. Mayeur, Hashem M. Mourad, Irene J. Beyerlein, T-3; Nathan A. Mara, Center for Integrated Nanotechnologies; Shuh-Rong Chen, George T. Gray III, MST-8	
<b>Hot Spots in the Electronic Spectrum Renormalize the Electronic Band Structure of Actinides .....</b>	<b>134</b>
Tanmoy Das, Jian-Xin Zhu, Matthias J. Graf, T-4; Tomasz Durakiewicz, John J. Joyce, MPA-CMMS	
<b>An Efficient Kernel Polynomial Method for Calculating Transition Rates in Large-Scale Materials .....</b>	<b>136</b>
Chen Huang, Arthur F. Voter, Danny Perez, T-1	
<b>Atomic Interface Design of Nanocomposites with Superior Shock Tolerance .....</b>	<b>138</b>
Ruifeng Zhang, T-3; Timothy C. Germann, T-1; Jian Wang, MST-8; Irene J. Beyerlein, T-3	
<b>Development of Interface-Dislocation Dynamics Simulations for Nanoscale Metallic Composites.....</b>	<b>140</b>
Caizhi Zhou, CNLS/T-3; Jian Wang, MST-8; Irene J. Beyerlein, T-3	
<b>Jamming, Pattern Formation, and Dynamic Phases for Driven Dislocation Assemblies.....</b>	<b>142</b>
Caizhi Zhou, CNLS/T-3; Irene J. Beyerlein, T-3; Charles Reichhardt, T-4; Cynthia Olson-Reichhardt, T-1	



<b>Funding Acknowledgment Cross Reference .....</b>	<b>ix</b>
<b>Author Cross Reference .....</b>	<b>144</b>
<b>Organizational Abbreviations (for this publication only) .....</b>	<b>148</b>
<b>Glossary .....</b>	<b>149</b>

# Funding Acknowledgment Cross Reference

American Recovery and Reinvestment Act (ARRA) .....	113
Department of Energy	
National Nuclear Security Administration (NNSA)	
Defense Science Programs	
Office of Advanced Simulation and Computing (ASC) Program .....	17, 19, 25, 71, 73, 75, 77, 81, 83, 85, 103, 105, 131, 137, 141
Weapons Science Campaigns .....	13, 15, 67
Defense Nuclear Nonproliferation .....	67
Office of Energy Efficiency and Renewable Energy (EERE)	
Vehicle Technology Program .....	7, 9
Office of Nuclear Energy .....	67
Office of Advanced Modeling and Simulation	
Consortium for Advanced Simulation of Light Water Reactors (CASL) .....	5
Office of Science .....	59, 71
Office of Advanced Scientific Computing Research (ASCR) Program .....	37, 81
Applied Mathematics.....	11, 21
Computational Science Research and Partnership (SciDAC) .....	37, 105
Exascale Co-design Center for Materials in Extreme Environments (ExMatEx) .....	89, 129
Office of Biological and Environmental Research (BER) .....	39, 41, 51, 53, 55, 57, 59
Climate Science for Sustainable Energy Future Project .....	39
Earth Systems Modeling .....	41, 51, 53, 55
Regional and Global Climate Modeling (RCGM) Program .....	37, 41, 51, 53, 55, 57, 59
Office of Basic Energy Sciences (BES).....	135, 137
Center for Materials at Irradiation and Mechanical Extremes (CMIME).....	139
Center for Integrated Nanotechnologies (CINT) .....	133
Office of Nuclear Energy.....	105
Department of Energy (DOE)/Department of Defense (DoD) Joint Munitions Program (JMP) .....	63, 131
Department of Defense (DoD) .....	115
Department of Defense (DoD)/Department of Energy (DOE) Joint Munitions Program (JMP) .....	63, 131
Naval Research Laboratory.....	41
European Commission's 7th Framework Programme COMBINE Project .....	41
European Union Projects .....	119
Harvard University, William F. Milton Award .....	113
National Aeronautics and Space Administration (NASA) .....	27



National Institutes of Health (NIH) .....	87, 113, 119
National Science Foundation (NSF) .....	69, 83, 95
Swedish International Development Corporation Agency.....	119
Swedish Research Council Postdoctoral Fellowship .....	119
Lawrence Livermore National Laboratory	
Laboratory Directed Research and Development (LDRD) Program.....	107
Los Alamos National Laboratory	
Director's Postdoctoral Fellowship .....	29, 31, 43
National Security Education Center .....	33, 51, 83
Institutional Computing Program .....	25, 85
Laboratory Directed Research and Development (LDRD) Program.....	25, 27, 29, 31, 33, 45, 47, 49, 65, 79, 91, 93, 99, 101, 105, 109, 113, 115, 117, 121, 123, 125, 129, 133, 135, 143
LDRD Postdoctoral Fellowship .....	29, 43
Oak Ridge National Laboratory .....	33











## Applied Mathematics and Fluid Dynamics

LANL has an extraordinarily rich scientific tradition of applied and scholarly research contributions in the fields of mathematics, fluid dynamics, and particle transport. Our impressive breadth of research impact speaks to this history, but also to the present and future with the significant integration of world-class theory, modeling, computational science, and experimentation. The collection of articles in this section underscores the breadth and depth of these seminal capabilities: from a new high-performance simulation tool for modeling important physical processes within a nuclear reactor, new methods applied to the understanding of engine combustion, multiple advances in the understanding of turbulent flow through direct numerical simulations, to advanced new methods that enable more accurate and faster hydrodynamic simulations with a variety of complex geometric and mathematical constraints. This sampling of work represents a much larger corpus of scientific capabilities embodied in complex multi-physics simulations that run on the largest supercomputers in the world for a host of national security, energy, and scientific programs.



# High-Fidelity Simulations of Turbulent Flow In Nuclear Reactors

Mark A. Christon,  
Jozef Bakosi,  
Robert B. Lowrie,  
Lori Pritchett-Sheats, CCS-2;  
Robert R. Nourgaliev, Idaho  
National Laboratory

We performed high-resolution simulations of the turbulent flow that occurs in the core of a pressurized water nuclear reactor. Our goal is to understand the flow-induced vibration problem that results in wear and failure of the fuel rods, forcing full-plant shutdowns and costing utilities millions of dollars every year. The simulations are building a database that will also be used to develop engineering models for future operational use. Analysis of the flow statistics enables a better understanding of how and why the efficiency of cooling is reduced over time and how failure of the fuel rods can be delayed.

**D**espite the recent Fukushima accident in Japan, worldwide production of electricity using nuclear reactors is increasing with more than 60 reactors under construction in 14 countries. In the US, 80% of the failures in pressurized nuclear reactors, where thousands of hot irradiated fuel rods are cooled by fast-flowing water, are caused by grid-to-rod fretting (GTRF) – a flow-induced vibration problem that leads to wear and failure of the rods. Contact points between the spring and the fuel rod cladding (Fig. 1) cause wear, leading to fuel leakage. If more than a few rods fail in an assembly, plants must be shut down and maintenance costs increase. Typical reactor fuel cycles range from 18 to 24 months, and increasing the life span of reactor fuel significantly improves the economic viability of nuclear power generation.

To understand the root causes of GTRF-induced fuel leaks, we investigate the turbulent coolant flow in the core of nuclear reactors. Our ultimate goal is to accurately predict why, how, and when wear occurs and how it can be prevented or, at least, delayed. Together with industry and university partners, we are developing a computational tool to support the decision-making process in the life-extension of current power plants and to enable the design of safer and more productive next-generation plants.

To date, it has not been possible to fully characterize the reactor-scale GTRF problem involving the thermal-hydraulics of turbulent multi-phase boiling flow coupled to the dynamics of thousands of vibrating fuel rods. As a first step, we report on large-eddy simulations of single-phase flow using two simplified rod bundle configurations ( $3\times 3$  and  $5\times 5$ ), as shown in Figs. 2 and 3. The simulations explicitly resolve the large-scale motion of the turbulent flow field using first principles and rely on a neutrally dissipative monotonicity-preserving numerical technique to represent the unresolved scales. A small section of a single fuel rod and a spacer grid are shown in Fig. 2 (top), depicting one of the surface meshes used for the computations. The spacer grid supports the rods and its mixing

vanes stir the flow in order to enhance the heat transfer between the rod and the coolant. This  $3\times 3$  geometry was extracted from a  $17\times 17$  fuel-rod assembly found in a typical reactor. There is a large degree of symmetry in the fuel assembly, which makes this geometrical simplification a reasonable approximation. The coolant flow generally moves from upper right to lower left in Fig. 2 (top).

The computational tool that we use, Hydra-TH [1], is being developed by our team at LANL and has the capability to compute high Reynolds number flows in very complex geometries. The code has been exercised with up to 10,000 compute cores to date using the LANL Institutional Computing resources. It yields a detailed time-accurate description of the fluid dynamics of the GTRF problem. As an example, isosurfaces of the helicity are plotted in Figs. 2 (bottom) and 4 at a single time instant for the  $3\times 3$  and  $5\times 5$  rod-bundle geometries, respectively. Larger helicity corresponds to a more violent turbulent flow and more efficient heat transfer. Although the vortex structures enhance the efficiency of the cooling, they also result in pressure fluctuations on the surface of the fuel rods, thus causing vibrations. The vibrations cause the fuel-rod supports to wear into the protective cladding on the rod surface, which, given enough time, may expose the internal nuclear fuel and result in a fuel leak. Experimental data, provided by Texas A&M University, have been used to validate the statistics of the simulated flow field for the  $5\times 5$  rod-bundle. For more details on the fluid dynamics simulations see Christon et al. [2] and Bakosi et al. [3].

We extract the forces acting on the rods in several downstream segments as time-histories, spectra, and integrated statistics from the fluid dynamics simulation. These data are then used to drive a structural dynamics code, VITRAN (Vibration TRansient Analysis–Nonlinear), developed by industry partners, to compute wear. VITRAN uses a modal dynamics approach and incorporates models for grid-to-rod gaps, fuel-rod cladding creep-down, spring relaxation, and grid growth. Understanding



Fig. 1. Fuel rod assembly.

#### Funding Acknowledgments

DOE, NNSA, Office of Nuclear Energy, Office of Advanced Modeling and Simulation, Consortium for Advanced Simulation of Light Water Reactors

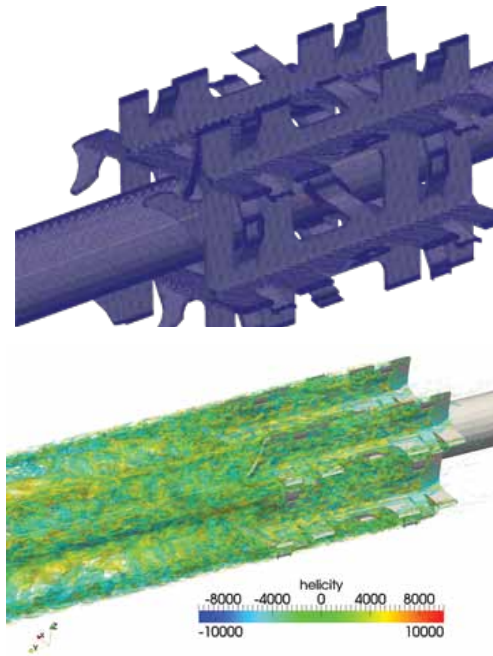


Fig. 2. (top) Surface mesh of small section of a single fuel-rod grid spacer with mixing vanes for the 3x3 rod bundle geometry. Approximate volume-cell count: 47 million. (bottom) Isosurface snapshots of the helicity for 3x3 rod bundle problem, approximate cell count: 47 million.

the detailed interactions of the fluid and structure is critical to mitigating the issues surrounding the GTRF problem.

The high-resolution fluid dynamics simulation data are also the basis for understanding the resolution requirements for future simulations and for developing lower-fidelity statistical turbulence models. These engineering models, specifically developed for the GTRF problem, can then be used operationally in an industrial setting to analyze and assess future rod-bundle designs, such as new spacer geometries, ultimately resulting in safer and higher-energy-throughput next-generation plants.

Future work on GTRF will focus on coupling the structural response of the fuel rods at different dynamic levels of approximation (e.g., one-way, two-way), along with coupling different wear models developed by collaborators in the Consortium for Advanced Simulation of Light Water Reactors (CASL) project, which also includes research staff from the MST and T divisions at LANL. The largest mesh we have run to date for the 3x3 and 5x5 problems have approximately 47 and 96 million computational cells, respectively. In order to adequately resolve the turbulent flow features and the heat transfer along the turbulent boundary layers, we believe that meshes of 100 million to 1 billion elements may be required, depending on plant operating conditions. To incorporate the effects of boiling, multiphase flow models are also being developed and implemented in the Hydra software toolkit.

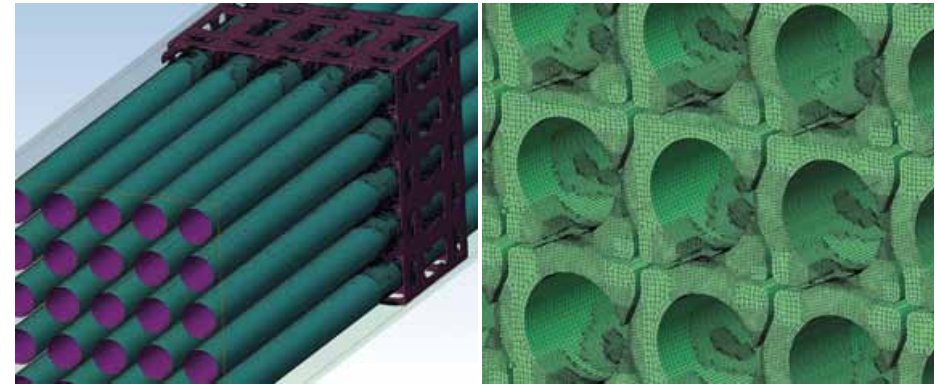


Fig. 3. (left) Surface mesh of spacer grid and fuel rods for the 5x5 rod bundle geometry. Approximate volume-cell count: 14 million. (right) Volume mesh around spacer grid and rods near the trailing edge of the mixing vanes.

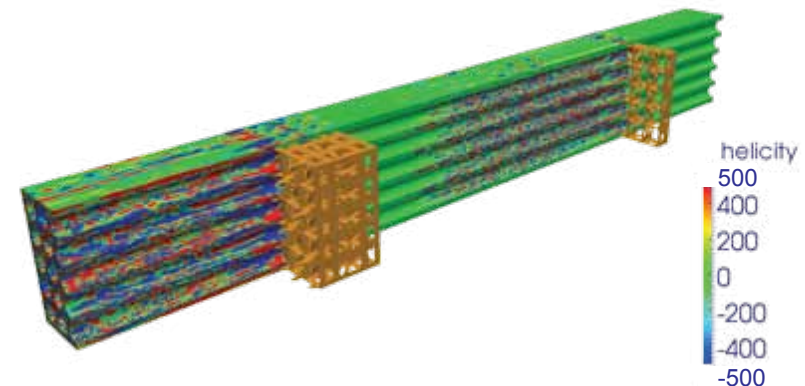


Fig. 4. Snapshots of helicity for the 5x5 rod bundle problem using an approximate cell count of 96 million.

[1] Christon, M.A., "Hydra-TH Theory Manual," LA-UR 11-05387 (2011).

[2] Christon, M.A. et al., "Initial Assessment of Hydra-TH Code on GTRF Problems," LA-UR 11-07034 (2011).

[3] Bakosi, J. et al., "GTRF Calculations Using Hydra-TH," LA-UR 12-24526 (2012).

# Modeling Fluid Flow in Domains Containing Moving Interfaces

David B. Carrington, T-3;  
Dominic A. Muñoz,  
Juan C. Heinrich, University  
of New Mexico

Numerical simulations of fluid flow in domains containing moving rigid objects or boundaries are still challenging when meshes have to be adapted and periodically regenerated. A numerical strategy that falls into the general category of Arbitrary Lagrangian–Eulerian (ALE) methods is being developed. The method is based on a fixed-mesh that is locally modified both in space and time to describe the moving interfaces that are allowed to move independently of the mesh. This results in a fully robust formulation capable of calculating in irregular meshes that contain moving devices of complex geometry and free of mesh entanglement problems. The present work constitutes the first stage in the development of a 3D model to interface with the new KIVA-hpFE simulator. The method's accuracy has been assessed in 2D using a case that has an analytical solution.

The design of internal combustion engines presents significant challenges to the optimization of shape, size, efficiency, power output, environmental impact, etc. Numerical simulations have provided an excellent tool for analysis prior to prototype building and have been used in the design of internal combustion engines for some time now [1,2]. The numerical models used to simulate flow in domains that physically change with time are for the most part based on ALE methods [3,4]. These methods, based on finite differences or finite elements formulations, are combined with moving mesh schemes in which the mesh is deformed or regenerated as the domain evolves to adapt to the changing geometry [5,6].

When the meshes are adapted to fit the evolving geometry they become degraded to the point where they must be regenerated, an expensive computational process, and they often become unusable, thus requiring the program operator's intervention. Our work is aimed at eliminating these problems by implementing an ALE method based on the use of a fixed, structured or unstructured mesh that covers the complete (or maximum) domain occupied by the fluid at any time in the simulation and that remains fixed throughout the calculation. The moving interfaces are described using sets of marker points that define the different moving bodies or boundaries. The marker points can move freely over the basic mesh with a velocity that may be prescribed or be part of the calculated solution [7]. Figure 1 illustrates a mesh and interfaces configuration. At each time step in the calculation the elements intersected by one of the moving interfaces are subdivided to fit the boundary with a piecewise linear curve such that the computational nodes always remain on the elements sides. The modified mesh is used to calculate the flow in the portion of the domain occupied by fluid only once for that time and interface position. At the next time step the interfaces are advanced, the new intersections with the mesh are calculated, and a new local

adaptation performed. Once the moving liquid-solid interface has gone through the stationary element, the element recovers its original form. Therefore, the mesh adaptation is performed only in those elements intersected by an interface and is local both in space and time. As a result, the method requires a minimal amount of interpolation and there are a fixed number of possible modifications to the intersected elements, three in the 2D case when quadrilateral elements are used and seven in the 3D case when hexahedrons are used. The situation is even simpler if the model is based on triangles in 2D and tetrahedrons in 3D, requiring only one kind of modification for 2D triangles and four in the case of four-node tetrahedrons. The use of this strategy eliminates the problem of maintaining the mesh quality and results in a robust formulation on arbitrary geometrical configurations.

At this time, the 2D four-node bilinear isoparametric element has been implemented to test the ideas, and the interface-moving algorithm has been combined with a first-order-in-time fractional step (projection) formulation [8] for incompressible flow. The accuracy of the model has been tested using an exact analytical solution to the case of 2D flow between two parallel plates separating with a prescribed velocity [9]. The error measures have been chosen by averaging over the computational nodes contained on a fixed a portion of the domain in order to obtain an estimate of the error at each time step, and then averaging those errors over a fixed-time interval. This results in one number representing the error in a computation for a specific spatial mesh. This process is repeated for two additional meshes, each refined so that the mesh size parameter is one half of the previous one, to estimate the convergence rate of the method as a function of mesh size. The results show the expected second-order convergence of the velocity and first-order convergence of the pressure as a function of the mesh size.

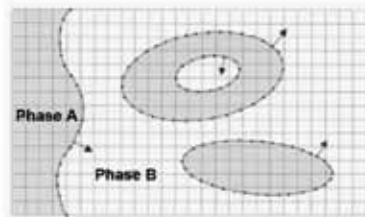


Fig. 1. Schematic of a rectangular domain discretized with a uniform mesh that contains a fluid phase (B) and three types of moving interfaces defined by marker points. The arrows indicate the positive vector normal to the interface.



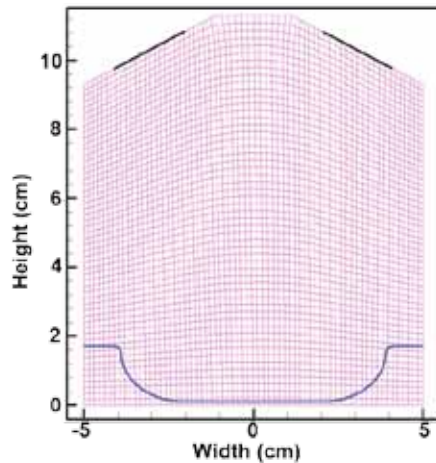


Fig. 2. Finite element mesh for the simulation of an engine piston chamber showing valve openings in black. The geometry and initial position of the top of the cylinder head is shown in blue at the bottom.

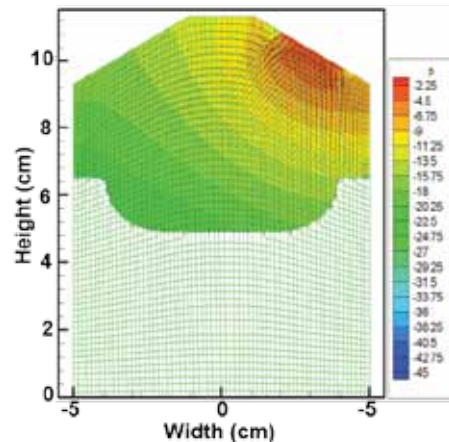


Fig. 3. Velocity and pressure after 3.5 seconds when the piston is moving back down and the inlet valve on the right-hand side is open. Maximum velocity is 4.4 cm/s.

An example that illustrates the capability of the method to model the interface's motion is given in Fig. 2. This calculation has been designed to show the effectiveness of the method to model the geometrical changes due to the moving interfaces; it does not include combustion, but it does consider 2D laminar incompressible flow at low Reynolds number. The figure shows a 2D idealization of a piston chamber and a cylinder head that moves in the y direction according to  $y(t) = y(0) + 2.5(1 + \sin(t - \pi/2))$ , so the stroke of the piston is 5 cm. Figures 3 and 4 show the position of the piston, the flow field, and the pressure field at two instants; in Fig. 3 it is starting to move down after 3.5 seconds of simulation; after reaching its maximum height, the left (outlet) valve is closed and fluid enters the region through the right (inlet) valve. In Fig. 4 at 6.55 seconds the piston is initiating upwards motion and the flow leaves the domain through the right valve. The simulation was performed in 13,100 time steps of equal size to reach 6.55 seconds. The total CPU time required for this run in a 2009 Dell Optiplex 960 PC with a 3-MHz, x86 Intel processor and 1.5 GB of physical memory is 3.79 minutes.

So far the present work has been restricted to laminar incompressible flow in two space dimensions in order to verify feasibility and accuracy of the method to model moving interfaces. The same methodology can be applied directly to high Reynolds number compressible and turbulent flows. The extension of the method to three space dimensions is currently under development.

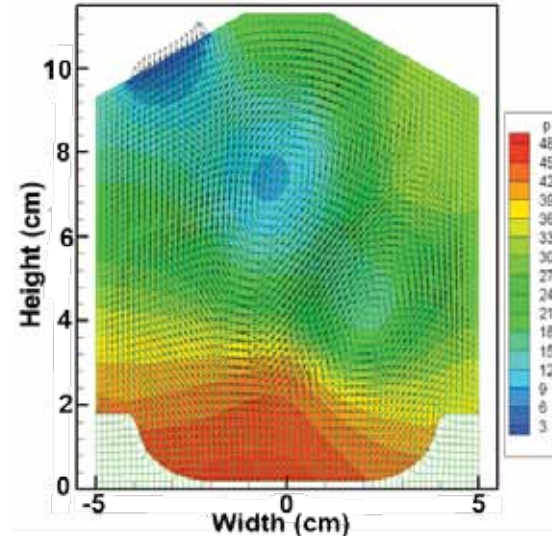


Fig. 4. Velocity and pressure after 6.55 seconds, the piston has just reversed its motion and is moving up. The inlet valve is closed and the outlet valve on the left is open. Maximum velocity is 7.7 cm/s.

- [1] Torres, D.J. and M.F. Trujillo, *J Comp Phys* **219**, 943 (2006).
- [2] Trescher, D., *Arch Comput Meth Eng* **15**, 67 (2008).
- [3] Donea, J. et al., *Comput Meth Appl Mech Eng* **33**, 689 (1982).
- [4] Farhat, C. and P. Geuzaine, *Comput Meth Appl Mech Eng* **193**, 4073 (2004).
- [5] Saksono, P.H. et al., *Int J Numer Meth Eng* **71**, 1009 (2007).
- [6] Löhner, R. and J.R. Cebal, *Arch Comput Meth Eng* **14**, 279 (2007).
- [7] Zhao, P. and J.C. Heinrich, *J Comp Phys* **173**, 765 (2001).
- [8] Quartapelle, L., *Numerical Solution of the Incompressible Navier-Stokes Equations*, Birkhäuser, Basel (1993).
- [9] Drazin, P. and N. Riley, "The Navier-Stokes Equations: a Classification of Flows and Exact Solutions," *London Mathematical Society Lecture Series* **344**, Cambridge University Press, Cambridge (2006).

# Adaptive Finite Elements with Local ALE for Modeling Turbulent Reactive Flow in Engines with Injection

David B. Carrington, T-3;  
Xiuling Wang, Purdue  
University;  
Juan C. Heinrich, University  
of New Mexico;  
Darrell W. Pepper, University  
of Nevada

LANL and collaborators are facilitating the effort for predictive internal combustion engine modeling. We are developing an  $hp$ -adaptive finite element methods (FEM) software that provides a high degree of accuracy and solution robustness. This FEM projection method, using a predictor-corrector scheme (PCS), has excellent capability over all flow regimes. The system is applicable to Newtonian and non-Newtonian fluids, turbulent reactive flows with sprays, and using a new local Arbitrary Lagrange-Eulerian (ALE) method for immersed moving parts.

These algorithms are capable of representing the physics within an engine. Theoretical Division resources and its collaborators provide software that many designers and researchers can use directly for engine simulation or may alter using their own models. KIVA software is in use worldwide by hundreds of universities and dozens of corporations involved in research, engine design, and manufacturing. We are working to deliver a more predictive capability to help in the understanding of combustion engines, thereby providing for greater efficiency and lower unwanted emissions.

LANL is continuing to develop advanced numerical methods that can be used in combustion and propulsion. We provide software that others may use directly or that they may alter with various models such as sophisticated chemical kinetics, different turbulent closure methods, or different fuel injection systems.

Current users are worldwide, from individuals and small research institutions to large corporations. We expect to deliver a more predictive and robust modeling capability. When put to use by researchers and corporations, we hope to better understand the complicated chemistry and physics associated with internal combustion engines. Greater efficiency is expected to result, helping to meet the newest US efficiency and pollution requirements for 2015, the aggressive requirements for 2020, and those in the future.

When considering the development of algorithms and the significant effort involved to produce reliable software, it is often best to create algorithms that are more accurate

at a given resolution only where and when it is required. We began developing a new KIVA engine/combustion code with this idea in mind [1]. This new construction is a Galerkin FEM approach that utilizes

conservative momentum, species, and energy transport. Our system uses a Petrov-Galerkin (P-G) and coupled-pressure stabilization [2].

A projection method is combined with higher-order polynomial approximation for model-dependent physical variables ( $p$ -adaptive) along with grid enrichment (locally higher grid resolution— $h$ -adaptive). Overset grids are used for actuated and immersed moving parts to provide more accurate and robust solutions in the next generation of KIVA. The scheme is particularly effective for complex domains such as engines.

The  $hp$ -adaptive FEM is, at a minimum, second-order accurate in space and third-order for advection terms, but becomes higher order where required as prescribed by the adaptive procedures [2]. The  $hp$ -adaptive method employs hierarchical basis functions, constructed on the fly as determined by a stress-error measure [3].

The  $h$ -adaptive method, along with a conservative P-G upwinding technique, accurately captures shocks. Figure 1 shows viscous supersonic Mach 2.25 occurring over an  $18^\circ$  compression ramp with the shock captured using adaptation. The velocity compares to experimental data as shown in Fig. 2 [4]. Differences are related to the  $k-\omega$  turbulence closure model—other models can capture the boundary layer more precisely but are more costly and can be less generic. The recirculation zone shown in Fig. 1(a) is an affect of the adverse pressure gradient developing in the boundary layer at the incline and is in agreement with the experimental data. The data only shows absolute speed, not direction. Taking this into account, the velocity for the  $k-\omega$  model is

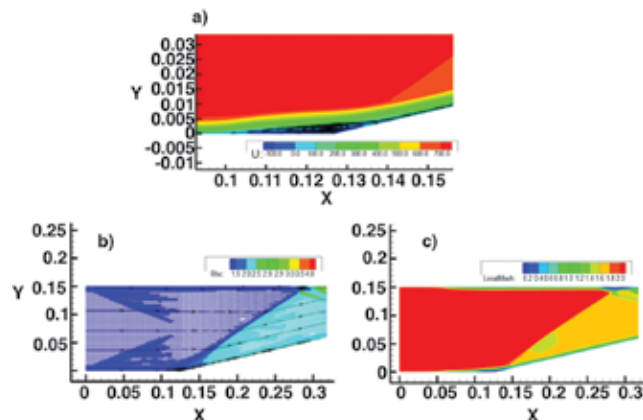


Fig. 1. Mach 2.25 steady-state flow properties for 2D supersonic viscous flow through a  $18^\circ$  compression ramp. (a) Recirculation, shock separation (distance in meters), (b) Density, (c) Local mach number.

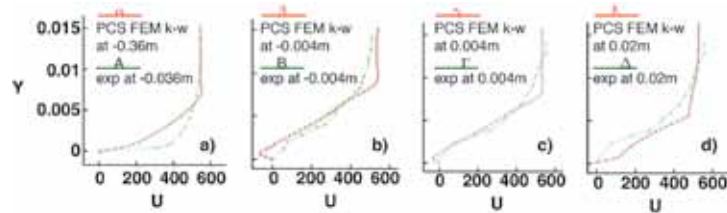


Fig. 2.  $U$  (mean velocity) in bottom boundary layer using  $k-\omega$  model. Comparison to data at two locations: Upstream(-) and downstream(+) of the ramp (a)  $-0.004\text{m}$ , (b)  $+0.004\text{m}$ , (c)  $+0.004\text{m}$ , and (d)  $+0.036\text{m}$ .

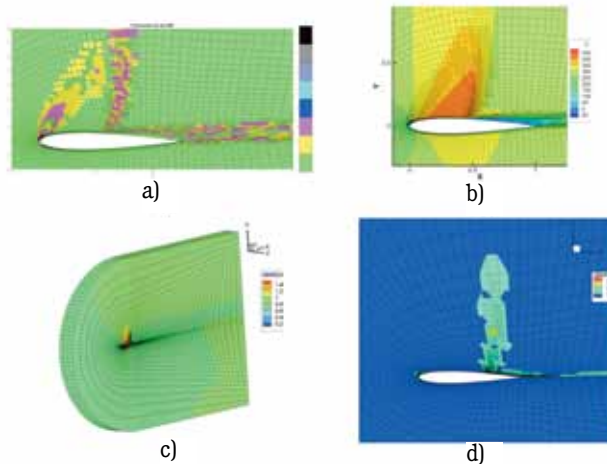


Fig. 3. Transonic flow over an NACA 0012 airfoil at  $4^\circ$  AOA, using  $hp$ -adaptive PCS FEM: (a) 2D  $hp$  grid with colors showing order of approximation. (b) 2D  $U$  component of velocity. (c) 3D Mach numbers. (d) 3D  $hp$  grid with colors showing order of approximation.

closer to the actual data than indicated in the Fig. 2(a) and (b).

Figure 3 shows both 2D and 3D subsonic/transonic flow over a NACA 0012 airfoil that also agrees with known solutions and data. This system incorporates a method for the measurement of the error in the discretization, and adjusts the spatial accuracy to minimize the error or bring it under some specified amount while minimizing the total number of nodes or elements in the domain.

The FEM method, when coupled with the spray models, provides a more accurate representation of droplet interaction with the conveying fluid and walls compared with the finite volume method used in the original version of KIVA. Because the FEM method allows for a continuous representation of phase-space, grid-scale accuracy can be applied everywhere.

Problems with coarse grids influencing the spray are related only to the solution accuracy—the spatial representation of the spray model is therefore convergent. The KIVA multi-component spray model, a method based on the algorithm developed by Dukowicz [5] and expanded by Torres et al. [6] for break-up, agglomeration, and surface films, is being installed in the  $hp$ -FEM PCS solver. In Fig. 4, diesel fuel is shown injected into a duct having developed flow and an inlet speed of 25 m/s (Mach 0.061) with an inlet Reynolds number of

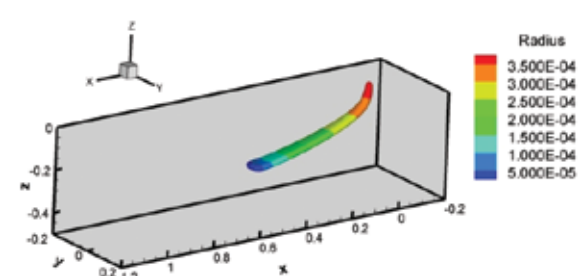


Fig. 4. Injection and spray modeling test case for PCS FEM. Diesel spray modeling: droplet breakup, agglomeration, and complete evaporation.

$\sim 1,204,000$ . The droplet radius is changing as a function of convective and diffusive heat and mass transfer, evaporating the droplets. Heat conduction within the droplet, collisions, and agglomeration occurs as the droplets are transported. Each parcel can have around 500 droplets (for 3D). The number of parcels/droplets is determined by the flow rate and injector size and is often in the neighborhood of 4000 parcels for about 200,000 droplets being represented.

Development of an  $hp$ -adaptive PCS FEM for all four regimes has been achieved. Compressible flow validation continues along with the implementation of the chemistry and spray models and new turbulence models, including large eddy simulation (LES) in wall-bounded domains. This projection method is a new solution algorithm for advancing the accuracy, robustness, and range of applicability of the KIVA combustion software suite. The system is higher-order both spatially and temporally yet provides a minimal amount of computational effort.

We have completed a local ALE technique for 2D that has second-order spatial convergence of error and will never tangle the grid [7]. The local ALE scheme uses overset grids for immersed parts described by their boundaries, which overlay the fluid grid. The moving parts within the fluid are not taken into account during the grid generation process. Hence, ports and cylinder portions of the grid are continuously represented. Because of this feature, the system allows CAD-to-grid in nearly a single step, providing nearly automatic grid generation.

- [1] Carrington, D.B., "A Characteristic-based Split  $hp$ -adaptive Finite Element Method for Combustion Modeling in KIVA- $hp$ FE", LA-UR 09-06527 (2009).
- [2] Carrington, D.B. et al., "A Predictor-Corrector Split Projection Method for Turbulent Reactive Flow," *Comput Therm Sci*, accepted (2012).
- [3] Wang, X. et al., "An Adaptive FEM Model for Unsteady Turbulent Convective Flow Over a Backward-facing Step," *Comput Therm Sci*, (2009).
- [4] Vallet, I., *Int J Numer Meth Fluids* **56**, 525 (2008).
- [5] Dukowicz, J.K., *J Comput Phys* **35**(2), 229 (1980).
- [6] Torres, D.J. et al., *Atomization and Sprays* **13**(2-3), 131 (2003).
- [7] Carrington, D.B. et al., "A Local ALE for Flow Calculations in Physical Domains Containing Moving Interfaces," *Adv Comput Fluid Dynam*, submitted (2012).



# M-Adaptation for Acoustic Wave Equation in 3D

Vitaliy Gyrya,  
Konstantin Lipnikov, T-5

Numerical modeling of wave propagation is essential for a large number of applied problems in acoustics, elasticity, and electromagnetics. The acoustic equation is one of the simplest examples of equation modeling wave propagation. For long integration times, the dominant contributions to an error in the solution come from such numerical artifacts as numerical dispersion and numerical anisotropy.

**N**umerical dispersion is the phenomenon in which the propagation velocity of the wave in the numerical scheme depends on its wavelength, while in the continuum problem there is no such dependence. Typically, the effect of the *numerical dispersion* is greater on under-resolved waves with ten or fewer points per wavelength, making them travel slower than in the physical problem. As a consequence, the wave does not simply arrive at a wrong time (which could be compensated for by time rescaling), but also has a highly distorted profile. *Numerical anisotropy* is the dependence of the numerical velocity of the wave on its orientation with respect to the mesh. We developed an m-adaptation technique for the acoustic wave equation in 2D on rectangular meshes. We identified the optimal numerical schemes in a rich family of second-order Mimetic Finite Difference (MFD) schemes that are fourth-order accurate for the numerical dispersion. On square meshes these schemes could be further optimized to be sixth-order accurate for numerical anisotropy. We refined the m-adaptation technique to be used in 3D on cuboid meshes. We identified fourth-order accurate schemes in the family of second-order MFD schemes. The resulting schemes have a nearly optimal time-step stability condition within the family.

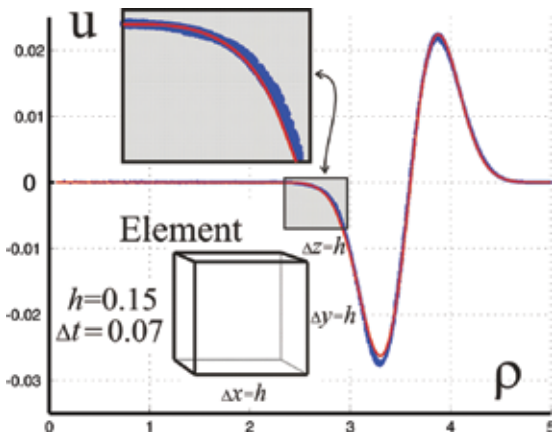


Fig. 1. Relative error in the numerical speed of the wave  $ch$  as a function of resolution parameter  $\kappa h$  (number of points per wavelength  $N = \frac{2\pi}{\kappa h}$ ) for

various directions  $\kappa$  of the wave  $\exp(\vec{\kappa} \cdot \vec{x} - ct / \kappa)$ . The m-adaptation on cuboid mesh (right) has the same fourth-order convergence rate as the modified quadrature method on cubic mesh (left).

The original and the semi-discrete forms of the acoustic wave equation in the time domain formulation are

$$u_{tt} = c \Delta u \text{ and } Mu_{tt} = Au \quad (1)$$

where the mass and stiffness matrices  $M$  and  $A$  are assembled from elemental matrices  $M_E$  and  $A_E$ , and where  $c$  is the wave-speed. Since the mass matrix  $M$  has to be inverted on every time step, the explicit time discretization of equation (1) is computationally efficient only when the inverse  $M^{-1}$  is easy to compute. One of the approaches is to replace the mass matrix  $M$  with a diagonal matrix  $D$  by lumping nondiagonal entries

to the diagonal. This does not change the order of the numerical scheme, but may lead to an undesirable increase of numerical dispersion.

Another approach [1] is to replace the inverse  $M^{-1}$  with the product  $D^{-1}MD^{-1}$ , where the inverse is taken only for the diagonal matrix  $D$ . Similar to lumping, this approach does not change the order of the numerical scheme. One can modify the stiffness and the mass matrices  $A$  and  $M$  using modified quadrature rules as is done in [2]. On square and cubic meshes, this approach produces schemes that are fourth-order accurate for numerical dispersion; however, this approach fails to do so in the more challenging case of rectangular and cuboid meshes.

Our approach has some similarities with [2] but is significantly more general. In fact, the schemes produced by [2] are a subset of the schemes analyzed in our approach, dubbed m-adaptation. We consider a parameterized MFD family of numerical schemes from which we select a member with the smallest numerical dispersion and anisotropy. The parameters in the MFD family appear through the elemental mass and stiffness matrices  $M_E^{MFD}$  and  $A_E^{MFD}$ , respectively. In 3D, the number of parameters is significantly larger than in 2D, so the techniques used in 2D are no longer tractable. For example, in 2D on a rectangular mesh the elemental matrices  $M_E^{MFD}$  and  $A_E^{MFD}$  depend on two parameters and one parameter  $\zeta$ , respectively. In 3D on cuboid meshes the elemental matrices  $M_E^{MFD}$  and  $A_E^{MFD}$  depend on 28 and 10 parameters, respectively.

The MFD family parameterized by  $(m_i, \zeta)$  contains a large number of known methods as special cases, for example, standard Finite Difference (FD), rotated FD, weighted combination of standard and rotated FD, Finite Element (FE) with lumped mass matrix, and modified quadrature method [2]. Moreover, compared with the later method, the MFD family is richer—containing 36 more parameters.

For the acoustic wave equation in 2D we selected the optimal parameters  $(m_1, m_2, \zeta)$  based on the von-Neumann analysis. In 3D, due to a much larger number of parameters, this approach was no longer tractable. We replaced this approach with another one, where we cancel the errors

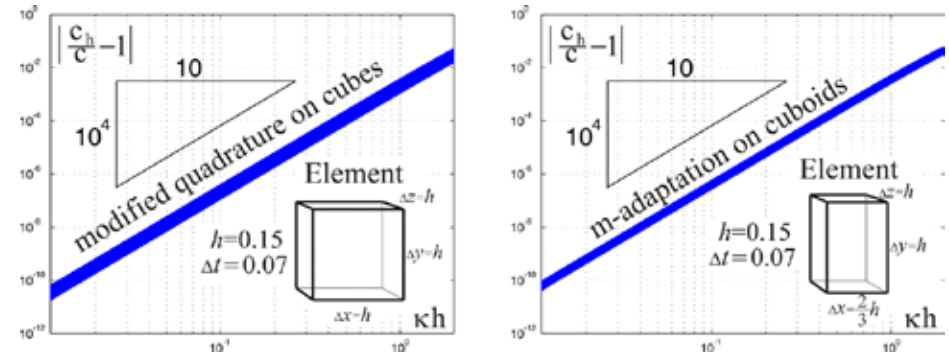


coming from the spatial and temporal discretizations for plane polynomial waves  $(\vec{\kappa} \cdot \vec{x} - \omega \tau)^p$  of degrees  $p = 1, \dots, 4$  for all possible directions and magnitudes of vector  $\vec{\kappa}$ . The seemingly infinite set of conditions for canceling the two errors can be condensed into a system of 21 equations that depend bilinearly on the elemental mass and stiffness matrices (thus on the parameters). The current state of the art for solutions for systems of bilinear equations does not allow for writing the solution for the system in an explicit form. Therefore, we have to rely on numerical solution of the system.

In addition to satisfying the above-mentioned 21 conditions, the set of 38 parameters has to yield positive definite mass and stiffness matrices. Moreover, the largest time step for which the scheme is numerically stable is inversely proportional to the largest eigenvalues of the matrices. Thus, we have to control both the largest and smallest eigenvalues of the mass and stiffness matrices. We identified and implemented an iterative numerical procedure for the solution of the above system subject to the optimization of eigenvalues. Based on this procedure, we found numerical schemes among second-order accurate MFD schemes that are fourth-order accurate for numerical dispersion both on cubic and non-cubic cuboid meshes.

We tested the optimized schemes using the dispersion relation, which in 3D has a similar form to the one we obtained in 2D for rectangular meshes. Once presented in logarithmic scale for the relative numerical error, it clearly shows fourth-order accuracy for the numerical dispersion (Fig. 1). This is the same accuracy as obtained by the modified quadrature scheme [2], but now it is achieved on general cuboid meshes. Moreover, on cubic meshes—although we obtained the same fourth-order accurate schemes as the modified quadrature schemes—our schemes had a stable time step that was larger by at least 10%.

As another test, we simulated a radially symmetric wave spreading from the origin, starting with Gaussian displacement and zero initial velocity. The radial symmetry



tests the numerical anisotropy, while the Gaussian profile (containing all wave frequencies) tests the numerical dispersion. The test shows that the optimized scheme on cuboid mesh  $\Delta x = .1$ ,  $\Delta y = \Delta z = .15$ , with  $\Delta t = .07$  has comparable dispersion to that of the modified quadrature scheme on a cubic mesh  $\Delta x = \Delta y = \Delta z = .15$  with  $\Delta t = .07$  and both produce very little dispersion for a mean wavelength of the Gaussian corresponding to 12 points per wavelength ( $\kappa h \approx 0.5$ ).

In the future, we plan to develop the m-adaptation technique for higher order schemes on general meshes and for elastic wave equations. The advantages of m-adaptation are that at a cost of some preprocessing one finds a fourth-order accurate scheme that has the complexity of a second-order one, requiring no matrix inversion during time step, therefore making it very efficient and accurate at the same time.

Fig. 2. Displacement as a function of the distance from the origin at time  $T = 3.6$  obtained using the modified quadrature method on cubic mesh (left) and the m-adaptation method on cuboid mesh (right) for a Gaussian initial displacement data.

[1] Gyrya, V. and K. Lipnikov, "M-Adaptation Method for Acoustic Wave Equation on Square Meshes," LA-UR 12-10047; *J Comput Acoust* **20**(4), 1250022:1-23 (2012).

[2] Guddati, M.N. and B. Yue, *Comput Meth Appl Mech Eng* **193**, 275 (2004).

# Late-Time Quadratic Growth in Single-Mode Rayleigh-Taylor Instability

Daniel Livescu,  
Tie Wei, CCS-2

The growth of the single-mode Rayleigh-Taylor instability has been investigated using Direct Numerical Simulations (DNS) [1]. The results show that, at long times and sufficiently high Reynolds numbers, the bubble acceleration becomes stationary, indicating mean quadratic growth. This is contrary to the general belief that single-mode RTI reaches constant bubble velocity at long times. A new stage, chaotic development, was found at sufficiently high Re values. During this stage, the instability experiences seemingly random acceleration and deceleration phases, due to complex vortical motions, with strong dependence on the initial perturbation shape. Nevertheless, the mean acceleration of the bubble front becomes constant, with little influence from the initial shape of the interface. As Re is lowered to small values, the later instability stages are subsequently no longer reached. Therefore, the results suggest a minimum Reynolds number and development time necessary to achieve all stages of single-mode RTI development, requirements that were not satisfied in the previous studies.

Rayleigh-Taylor instability (RTI) is an interfacial instability that occurs when a high-density fluid is accelerated or supported against gravity by a low-density fluid. This instability is of fundamental importance in a multitude of applications ranging from fluidized beds, oceans and atmosphere, to Inertial Confinement Fusion (ICF) and supernovae [2]. Thus, the loss of target performance in ICF is associated with the development of RTI. Better knowledge about the growth of the instability and its dependence on initial conditions can help more accurately predict important practical problems. Using DNS, we have systematically studied the development and dependence on initial conditions of single-mode RTI [1]. Besides its own interest, single-mode RTI has also been used as a building block for the study of multimode RTI development. Despite its apparent simplicity, single-mode RTI is still not well understood and continues to be the focus of research in experimental (with both 2D and 3D perturbations), numerical, and theoretical studies.

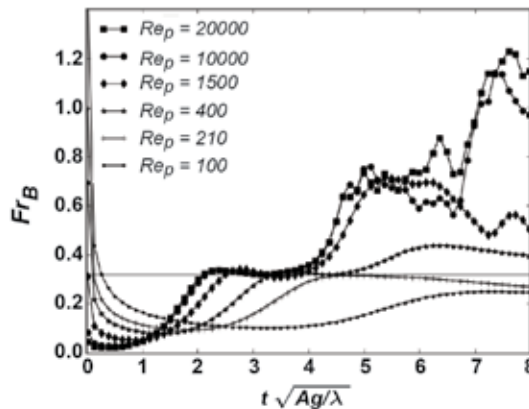


Fig. 1. The effect of the Reynolds number on the normalized bubble front velocity,  $Fr_B$ .

All simulations have been performed with the CFDNS code [3]. We have carried out extensive resolution studies to ensure that the solution is converged and the flow symmetries are preserved. The simulation results also show excellent agreement with the Linear Stability Analysis (LST), the analytical prediction of Goncharov [4], and the experimental results of Waddell et al. [5].

The development of single-mode RTI is usually divided into a number of stages, depending on which physical effect dominates the instability growth. We found that this development and the transition (or lack

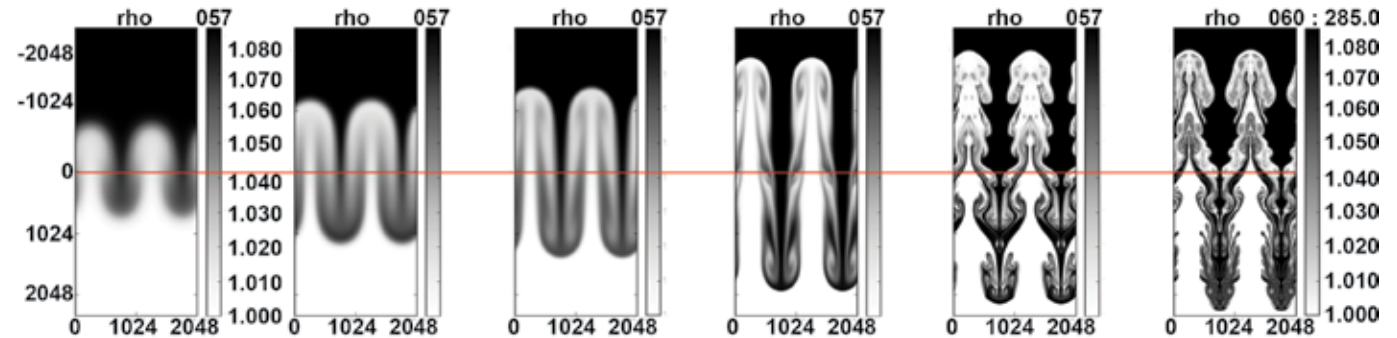
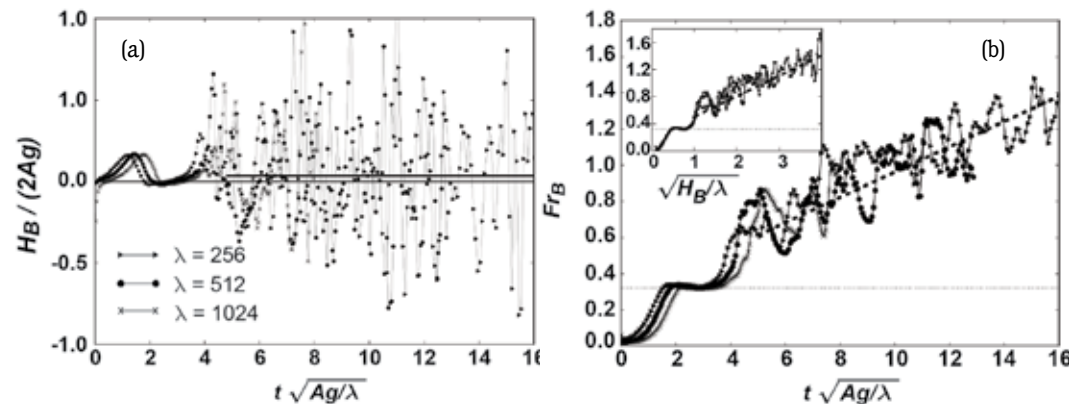
thereof) to the subsequent stages is strongly influenced by a perturbation Reynolds number, defined by  $Re_p = \lambda \sqrt{A/(A+1)\lambda g}$ . The full range of development stages can be obtained at high enough values of the Reynolds number. Until recently, it was believed that the bubble front approaches and maintains a constant “terminal velocity” at late times (i.e., linear bubble height growth,  $H_b \sim t$ ), based on the assumption that the flow remains potential near the tip of the bubble. Our results show that the flow does indeed reach such a regime (Fig. 1); however, while the vorticity is zero in this region due to symmetry conditions, the induced velocities due to vortical motions inside the bubble become strong enough to render such potential flow solutions inadequate at long times. Thus, at high enough Reynolds numbers, the Kelvin-Helmholtz instability generates vortices at the interface as the two fluids move in opposite directions. These vortices are first generated near the initial interface, then break into vortex pairs (in the 2D case) or rings (in the 3D case), that self-propagate towards the tips of the bubbles and spikes. The induced velocities generated by the vortex pairs or rings add bursts of acceleration to the instability growth, until these reach the tips of the bubbles and spikes, where they break into smaller structures. However, vortices are continuously being generated at the interface and the process continues. The resulting flow exhibits complex vortical interactions (Fig. 2). Although the flow should preserve the initial reflectional symmetries (i.e., the axes of the bubbles and spikes), the motions in between these symmetry lines become chaotic and the bubble acceleration becomes stationary, with non-zero mean. The instantaneous evolution of the layer shows strong sensitivity to the initial conditions; however, the mean acceleration has little influence from the initial perturbation shape. The vortical interactions, neglected in previous

Fig. 2. Density contours at late time for different Reynolds number simulations. From left to right,  $Re_p =$  (a) 100, (b) 210, (c) 400, (d) 1,500, (e) 10,000 and (f) 20,000.

studies, transform the constant velocity growth into mean quadratic growth at high enough Reynolds numbers (Fig. 3).

These results show that there is no fundamental difference between the single-mode and multimode RTI growth, as previously believed. In addition, the results reconcile the apparent contradiction between the previously believed constant velocity single-mode growth and recent results showing fast quadratic growth (with  $\alpha$  values larger than those routinely obtained in numerical simulations) for initial perturbations with a pronounced peak at  $k = 1$  [6]. Therefore, single-mode RTI results could be used to understand the growth of laterally confined RTI (when the  $k = 1$  mode dominates the spectrum).

Fig. 3. (a) Bubble front acceleration, and (b) bubble front velocity evolution at  $Re_p > 10,000$  showing mean quadratic growth.



number can be defined using the SGS (or numerical) viscosity. LES of single-mode RTI with coarse mesh may suffer from a low  $Re_p$  effect, due to an insufficient range of scales allowed by the mesh. Such an effect can be clearly seen in the Implicit Large Eddy Simulations (ILES) results of Ramaprabhu et al. [7] (see their Figs. 11 and 12), Glimm et al. [8] (their Fig. 1), Francois et al. [9], and Ramaprabhu et al. [10]. In those studies, the changing of meshes [7,8] or numerical methods [9] gave different growth rates, due to the implicit change in the effective perturbation Reynolds number. In coarse mesh (or lower effective  $Re_p$ ) simulations, the later instability stages were not observed. Even when the instability reached a growth stage beyond the constant velocity regime (e.g., [10]) the flow returned to bubble front velocities smaller than the potential flow result, presumably due to insufficiently high Reynolds numbers. In

addition, when small-scale vortical motions were generated and seemingly influenced the growth, there is a clear interference between the physical and numerical vorticity production mechanisms, indicated by the breaking of the symmetries that should be preserved by the flow. Due to the sensitivity of the instability growth to the vortical motions, this raises significant questions on the relevance of ILES techniques to the single-mode RTI.

[1] Wei, T. and D. Livescu, *Phys Rev E* **86**, 046405 (2012).

[2] Livescu, D., et al., "Numerical simulations of Two Fluid Turbulent Mixing at Large Density Ratio and Applications to the Rayleigh-Taylor Instability," *R Soc London Phil Trans A*, to appear (2012).

[3] Livescu, D. et al., "CFNDS 1.0," LA-CC-09-100 (2009).

[4] Goncharov, V.N., *Phys Rev Lett* **88**, 134502 (2002).

[5] Waddell, J.T. et al., *Phys Fluids* **13**, 1263 (2001).

[6] Livescu, D. et al., *J Phys Conf Ser* **318**, 082007 (2011).

[7] Ramaprabhu, P. et al., *Phys Rev E* **74**, 066308 (2006).

[8] Glimm, J. et al., *Acta Mathematica Applicatae Sinica* **18**, 1 (2002).

[9] Francois, M.M. et al., *Proceedings 10th Int Workshop Phys Compressible Turbulent Mixing, IWPCTM10*, **77** (2007).

[10] Ramaprabhu, P. et al., *Phys Fluids* **24**, 074107 (2012).

# Direct Numerical Simulations of Rayleigh-Taylor Instability with Gravity Reversal

Daniel Livescu,  
Tie Wei, CCS-2

We present results from an extensive new set of Direct Numerical Simulations (DNS) of Rayleigh-Taylor instability (RTI). The set includes a suite of simulations with grid size of  $1024^2 \times 4608$  and Atwood number ranging from 0.04 to 0.9, in order to examine small departures from the Boussinesq approximation as well as large density ratio effects, and a high resolution simulation of grid size  $4096^2 \times 4032$  and Atwood number of 0.75, which is the largest instability simulation to date. After the layer width had developed substantially, additional branched simulations have been run under reversed and zero gravity conditions. These simulations represent unit problems for the variable acceleration case encountered in practical applications. While the bulk of the results is still being analyzed, here we focus on the modifications in the mixing layer structure and turbulence in response to the acceleration change [1].

Although RTI has been subjected to intense research over the last 50 years, until recently numerical studies have been restricted to coarse mesh calculations. On the other hand, it is notoriously difficult in laboratory experiments to accurately characterize and control the initial conditions and provide the detailed measurements needed for turbulence model development and validation. Thus, a large number of open questions remain unanswered about this instability and even first-order global quantities, such as the layer growth, are not completely understood and still give rise to intense debate [2,3]. Nevertheless, today's petascale computers allow fully resolved simulations of RTI at parameter ranges comparable to those attained in laboratory experiments, but providing, in carefully controlled initial and boundary conditions studies, much more information than physical experiments. These extremely high resolution simulations are enabling a look at the physics of turbulence and turbulent mixing in unprecedented detail, hopefully contributing to a significant advance in our understanding of these phenomena [4].

Although in many instances, for example in atmospheric or oceanic flows, the acceleration may be considered as constant, there are important practical applications where the driving acceleration changes in time and may even reverse sign. Some examples include Inertial Confinement Fusion (ICF), supernovae, and pulsating stars. RTI with complex acceleration history, including changes in sign, have been studied both experimentally and in coarse mesh numerical simulations; however, much less is known about the physics of the flow compared to the classical RTI. In addition, most engineering turbulence models cannot capture the change in the mixing layer behavior following gravity reversal. In order to better understand the variable acceleration effects on the instability

development and turbulence properties, we have proposed two unit problems: reversing or setting the acceleration to zero in the turbulent stage of the classical RTI [5]. The latter is related to the Richtmyer-Meshkov instability, when the two fluids are subjected to an impulsive acceleration, for example, due to a shock wave.

All simulations have been performed with the CFDNS code [6]. We have carried out extensive resolution studies to ensure that the solution is converged and all flow scales are accurately resolved. While the bulk of the results are still being analyzed, we present here, preliminary results from three simulations, with  $A = 0.75$ , representing classical, reversed, and zero gravity RTI cases.

The primary nondimensional parameter characterizing differential acceleration effects is the Atwood number,  $A = (\rho_h - \rho_l)/(\rho_h + \rho_l)$ , where  $\rho_l$ ,  $\rho_h$  are densities of the light and heavy fluids, respectively. The Atwood number ranges from 0 to 1. For air inter-penetrating helium, for which the density ratio is 7, the Atwood number is  $A = 0.75$ . For air and hydrogen,  $A = 0.85$ .

In contrast, the Boussinesq approximation corresponds to  $A \rightarrow 0$  and a value of 0.05 is usually taken to define this limit. To the best of our knowledge the new set of DNS addressed here is the first to consider  $A > 0.5$ . Yet the development of the instability and the mixing itself are fundamentally different at high and low  $A$  [2]. For example, our previous results in an idealized triply-periodic buoyancy-driven flow show that the mixing is asymmetric at large density ratios, with the pure light and heavy fluids mixing at different rates [7,8]. We refer to such flows as variable-density (VD) flows.

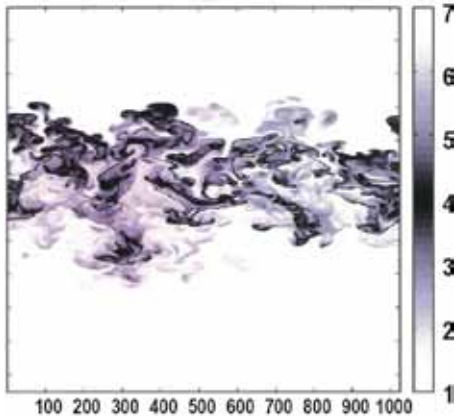


Fig. 1. 2D snapshot of the mixing layer from the forward gravity case at the reference time.



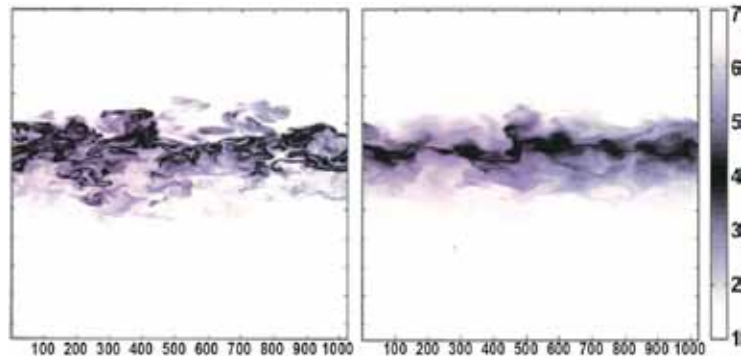
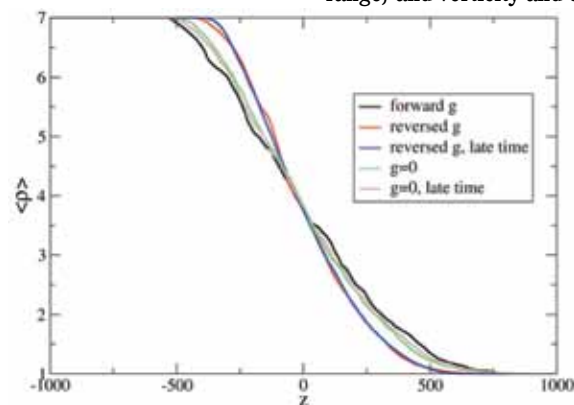


Fig. 2. 2D snapshot of the mixing layer from the reversed gravity case at the (left) reference time and (right) late time.

The change in the structure of the mixing layer after the gravity reversal can be clearly seen in Figs. 1 and 2. In the forward gravity case (Fig. 1), the interpenetration

of the two fluids is highly irregular, giving rise to local density inversions, similar to the previous lower A results [9]. These inversions are quickly removed as the buoyancy force reverses sign so that, at late times, the stratification becomes relatively uniform across the layer (Fig. 2). On the contrary, if the gravity is set to zero, the local structures are preserved (not shown here). As the layer continues to grow due to the inertia of the individual fluid parcels, the instability still develops and the constant density surfaces become even more corrugated. This is similar to the post re-shock evolution of the mixing layer in the Richtmyer-Meshkov instability.

Fig. 3. Mean density profiles corresponding to the three cases at the reference and late times.



In general, the spectral properties (in terms of fully developed spectra showing the emergence of an inertial range) and vorticity and density gradient alignments with the strain tensor eigenvectors remain similar, following the change in gravity, with the forward density case and with the usual canonical turbulent flows. However, for the reversed gravity case, the vector quantities enumerated above also acquire a strong directionality with the coordinate directions. In addition, gravity reversal efficiently mixes the largescales in the

inner region of the layer, so that the large-scale anisotropy decreases to zero and the stratification becomes uniform across the layer. However, the anomalous small-scale anisotropy associated with buoyancy-driven turbulence [2,9] is actually increased by gravity reversal.

On the other hand, the mean density profiles shown in Fig. 3 give little indication of the dramatic modifications in the underlying density field structure as the gravity changes. However, these modifications have a significant influence on the turbulence transport, rendering the popular gradient diffusion hypothesis, used in moment closures, not appropriate. For example, the mass flux is closed in two-equation models like  $k$ -L or  $k$ - $\epsilon$  as proportional to the mean density gradient. Figure 4 shows that the mass flux changes considerably (even reverses sign) following the gravity reversal, which clearly cannot be captured by a gradient diffusion hypothesis. Thus two-equation models cannot capture such flows; one needs a separate mass flux transport equation, as it is done, for example, in the BHR model [10].

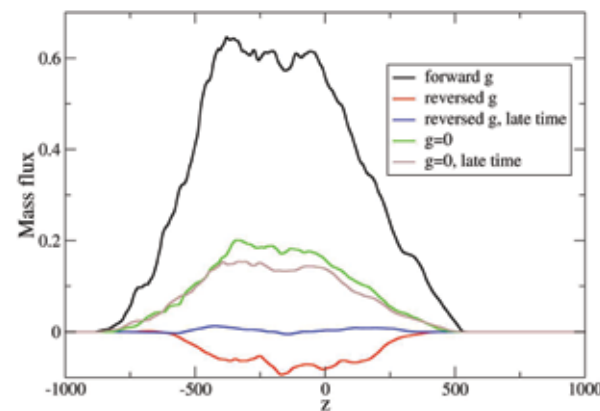


Fig. 4. Mass flux variation across the layer for the three cases at the reference and late times.

- [1] Livescu, D. and T. Wei, *Proceedings 7th International Conf Comput Fluid Dynam*, (ICCFD7), 2304 (2012).
- [2] Livescu D. et al., *Phys Scripta* **T142**, 014015 (2010).
- [3] Dimonte G. et al., *Phys Fluids* **16**, 1668 (2004).
- [4] Livescu, D. et al., "Numerical simulations of Two Fluid Turbulent Mixing at Large Density Ratio and Applications to the Rayleigh-Taylor Instability," *R. Soc London Phil Trans. A* (to appear).
- [5] Livescu, D. et al., *J Phys Conf Ser* **318**, 082007 (2011).
- [6] Livescu, D. et al., "CFDNS: A Computer Code for Direct Numerical Simulations of Turbulent Flows," LA-CC-09-100.
- [7] Livescu, D. and J. R. Ristorcelli, *J Fluid Mech* **591**, 43 (2007).
- [8] Livescu, D. and J. R. Ristorcelli, *J Fluid Mech* **605**, 145 (2008).
- [9] Livescu, D. et al., *J Turbul* **10**, 1 (2009).
- [10] Schwarzkopf, J. D. et al., *J Turbul* **12**, 1 (2011).

# Turbulence Model Validation for Complex Mixing Scenarios

Bertrand Rollin, CCS-2;  
Nicholas A. Denissen,  
Jon M. Reisner,  
Malcolm J. Andrews, XCP-4;  
Jimmy Fung, XCP-1

Computing requirements for simulating complex multi-physics flows are such that Reynolds averaged Navier-Stokes (RANS) models will remain the community's chosen workhorse for years to come. Within applications of interest to LANL, such as compressible mixing, the Besnard Harlow Rauen Zahn (BHR) RANS models are being utilized within Advanced Simulation and Computing (ASC) codes. Specifically, the present work examines the ability of the BHR-2 turbulence model to simulate Rayleigh-Taylor mixing and bulk interface motions of two fluids driven by gravity within a tilted rig experiment. We present here the results from simulations using the BHR-2 model and compare them to available experimental and Implicit Large Eddy Simulation (ILES) results. These comparisons are intended to demonstrate the utility of the BHR-2 model to accurately predict various aspects of compressible turbulent mixing.

**T**urbulence and turbulent mixing are ubiquitous in nature and engineering applications. Whether considering numerical weather prediction, combustion efficiency in a scramjet engine, or performance of an Inertial Confinement Fusion (ICF) capsule, the effects of turbulence cannot be disregarded. However, the multi-scale and 3D nature of turbulent flows is such that turbulence modeling, as opposed to direct numerical simulation of a given flow field, is the only practical approach when studying full-scale engineering problems. In particular, RANS models remain the standard tool for a wide variety of complex, multi-physics applications.

To address the range of applications of interest at LANL, the BHR family of compressible turbulent mix models were developed and then implemented in LANL ASC models such as xRage [1]. The present work uses one member of the BHR family, BHR-2, which solves transport equations for kinetic energy,  $K$ , turbulence length scale,  $L$ , turbulent mass-flux velocity,  $a_i$ , and density-specific volume correlation,  $b$ . BHR-2 has been demonstrated for canonical turbulent flows such as Kelvin-Helmholtz, Rayleigh-Taylor, and Richtmyer-Meshkov [4] where, on average, mixing is 1D. However, in many problems of interest, there can be significant curvature of the interface as well as dynamic interface movement. How BHR-2 handles these effects, both the influence of the turbulence on the bulk motion of the interface, and vice versa, is essential for the tilted-rig experiment shown in this paper, as well as within more complex applications of interest to LANL.

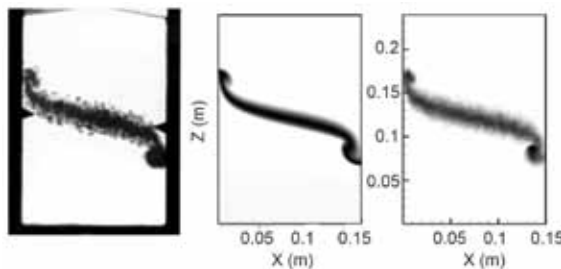


Fig. 1. Mixing region at  $t=45$  ms from experiment (left), a LANL hydrocode, FLAG, with BHR-2 modeling (middle), and ILES (right).

Tilted-rig experiments were originally run in the 1980s and early 1990s by Smeeton and Youngs [5], Andrews and Spalding [6,7], and Pttizyna et al. [8]. In these experiments, a tank filled with a light fluid above a heavy fluid is tilted a few degrees off the vertical causing a slanted interface. Upon acceleration of the tank by rocket engines and a resulting reversal of gravity, large-scale overturning motion quickly commences producing rising and falling plumes at the side edges of the tank and Rayleigh-Taylor-driven mixing at the center of the tank.

Figure 1 shows comparisons between the photograph of the actual experiment [5], the RANS, and ILES simulations at  $t = 45$  ms. Close inspection of this figure reveals that the RANS simulation is able to reproduce with good accuracy the bubble (right-hand-side fluid structure) and spike (left-hand-side fluid structure) penetration distance as well as their overall shape, but slightly under-predicts the mixing layer size. These comparisons reveal that the flow dynamics displayed by the ASC hydrocode coupled to the BHR-2 model is consistent with the flow dynamics observed in the experiments and ILES.

Further assessment of the model requires comparisons between the turbulence model variables as predicted by BHR-2 and inferred from the ILES. The first quantity of interest is the turbulence kinetic energy,  $K$ , as it is the most affected by the motion of the interface. Figure 2 shows the distribution of  $K$  as predicted by RANS modeling and ILES at  $t=45$  ms. Both kinetic energy contours show identical dominant features. Local maxima of  $K$  can clearly be identified at the tip of the bubble and spike. The widening and increase in intensity of  $K$  when moving toward the bubble side is well captured by the RANS simulation as well as the asymmetry of the distribution. Next, the turbulent mass-flux velocity—a key quantity in compressible turbulence as it is the primary production

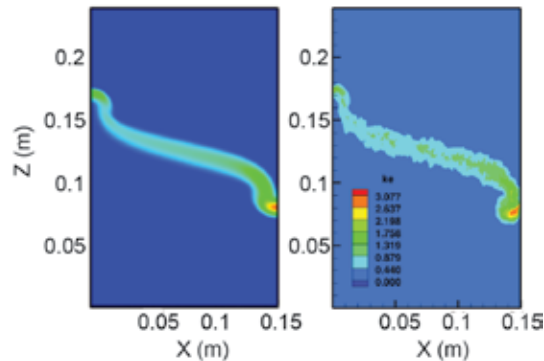


Fig. 2. Distribution of turbulent kinetic energy,  $K$ , as predicted by a LANL hydrocode, FLAG, with BHR-2 (left) and ILES (right) at  $t=45$  ms.

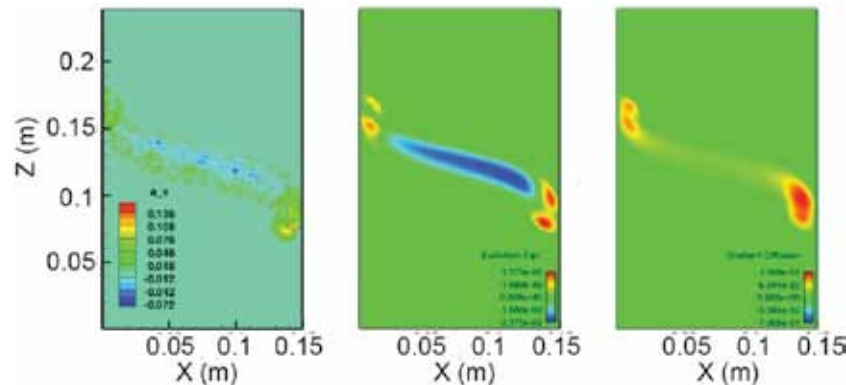


Fig. 3. Distribution of horizontal mass flux,  $a_x$ , as predicted by ILES (left), a LANL hydrocode, FLAG, with BHR-2 (middle), and a LANL hydrocode (FLAG) with a classic K/L (two-equation) turbulence model (right) at time  $t=45$  ms.

term for turbulent kinetic energy—is compared against corresponding results from a commonly used two-equation RANS model. As evident in Fig. 3, BHR-2 is able to predict the negative (counter-gradient) horizontal turbulent mass-flux velocity in the center of the mixing layer seen in ILES, contrary to the two-equation turbulence model. This finding suggests the better suitability of the four-equation BHR-2 model over two-equation models for handling complex mixing scenarios.

Key findings of the BHR-2 turbulent-mix model's validation for problems with dynamic interfaces have been presented. The BHR-2 model can efficiently predict a variety of turbulence quantities including turbulent kinetic energy, turbulent mass-flux and density-specific volume correlation (not shown). The model does not adversely affect the mean interface motion, and accurately captures differences in turbulence intensity along the moving interface. Detailed analysis of this work and future perspectives can be found in Denissen et al. [9] and Rollin et al. [10].

[1] Besnard, D. et al., *LANL Technical Report* "Turbulence Transport Equations for Variable-Density Turbulence and thier Relationship to Two-Field Models." LA-12303-MS (1992).

[2] Banerjee, A. et al., *Phys Rev E* **82**, 046309 (2010).

[3] Schwarzkopf, J.D. et al., *J Turbulence* **12**, N49 (2011).

[4] Stalsber-Zarling, K. and Gore, R.A., "The BHR2 Turbulence Model: Incompressible Isotropic Decay, Rayleigh-Taylor, Kelvin-Helmholtz and Homogeneous Variable Density Turbulence." LANL Technical Report LA-UR-11-04773 (2011).

[5] Smeeton, V.S. and D.L. Youngs, "Experimental Investigation of Turbulent Mixing by Rayleigh-Taylor Instability Part 3." AWE Report No. O 35/87 (1987).

[6] Andrews, M., *Ph.D. Thesis*, "Turbulent Mixing by Rayleigh-Taylor Instability." London University (1986).

[7] Andrews, M. and Spalding, D., *Phys Fluids A* **2**(6), 322-327 (1986).

[8] Pützyna N.V. et al., *Proceedings 4th IWPCTM*, 729-743 (1993).

[9] Denissen N.A. et al., *Proceedings 43rd AIAA Fluid Dynamics Conference and Exhibit*, in preparation (2013).

[10] Rollin, B. et al., *Proceeding IMECE 2012*, IMECE2012-93094 (2012).

# Towards a Next-Generation Reconnection-ALE Hydrocode

**Shiv K. Sambasivan, CCS-2;**  
**Mikhail J. Shashkov, XCP-4**

The focus of our research is the development of a new Reconnection-Arbitrary-Lagrangian-Eulerian (ReALE)-based hydrocode for modeling solids and fluids subject to extreme conditions. The ReALE formulation that we are developing is based on a physically motivated and mathematically rigorous construction of the underlying Lagrangian method and vector/tensor reconstruction and remapping algorithms. This work brings together many new concepts that in combination with contemporary cell-centered Lagrangian methods will produce a cutting-edge ReALE capability and define a new state-of-the-art. The proposed research and the resulting algorithms will be of immediate use in Eulerian, Lagrangian, and Arbitrary Lagrangian Eulerian (ALE) codes in several DOE programs and codes.

**M**odeling real materials with finite strength is of particular interest in applications such as Inertial Confinement Fusion (ICF), munition-target interactions, geological impact dynamics, shock processing of powders, and formation of shaped charges. These applications constitute a specific class of problems in which the hydrodynamic pressure realized is often much greater than the strength of the material, thereby leading to large deformation of the interacting media. Traditionally, the tools that have been used to solve large deformation transient problems have been termed hydrocodes.

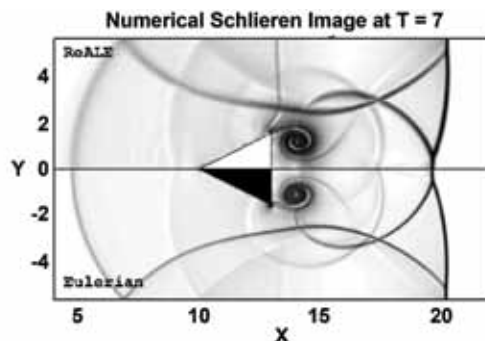
Over the past 20 years, there have been a plethora of significant hydrocode efforts based on Eulerian, Lagrangian, and ALE methods. In spite of their number, existing hydrocodes are known to exhibit multiple numerical, physical, and mathematical inconsistencies. For instance, despite the recent advancement with the Eulerian formulations, the Lagrangian framework has been widely preferred for many applications due to its ability to follow the material response. Furthermore, since many of the constitutive relations demand detailed knowledge of the

material motion, Lagrangian formulations are considered to be the natural choice. On the other hand, Lagrangian formulation-based ALE methods require frequent surgical procedures to prevent mesh from tangling, particularly for applications with strong vortices. Additional pathologies stem from the use of low-order staggered-grid discretizations, ad hoc artificial viscosities, unconstrained mesh instabilities, and poor pressure-gradient approximations on distorted grids.

The primary goal of this research is to develop a numerical paradigm that is devoid of the pathologies and yet faithfully represents the material behavior. In particular, we are developing a new ReALE-based hydrocode [1] for solving large deformation problems in solids and fluids. As the name indicates,

ReALE permits the connectivity of the mesh to dynamically adapt to evolving features in the solution field. The reconnection part of ReALE significantly alleviates the computational encumbrance enjoined with mesh untangling procedures. Furthermore, applications with intrinsic strong vortices can be captured and treated adequately. At the heart of our algorithm lies a high-order cell-centered discretization scheme [2] that neither requires explicit artificial viscosity terms nor demands ad hoc mesh stabilization methods. The underlying construction of the scheme is generic in that advanced hypo/hyperelastic constitutive theories can be easily incorporated. Built upon this emerging cell-centered scheme are the physically motivated and mathematically consistent, frame invariant and symmetry preserving, vector/tensor reconstruction, limiting, and remapping techniques. Consequently, we are aiming to remove numerical and physical inconsistencies at a more fundamental level. Together with these significant improvements, we are establishing the next generation ReALE hydrocode as an alternative to Eulerian and ALE methods.

Our preliminary investigation demonstrates an impressive outcome for applications involving fluid and solid media subject to extreme conditions. The capability of the formulation in capturing dominant features in the solution field is evinced in Figs 1, 2, 3, and 4. In these figures, the new ReALE formulation-based solution is compared against well-established Eulerian methods [3] for problems that are inherently Eulerian in nature. In Fig. 1, we demonstrate the ability to capture strong vortical structures in the flow field. A Mach 1.3 shock impacting a wedge-shaped rigid obstruction generates a complex shock-vortical system. Comparison with Eulerian simulation indicates remarkable agreement between the two calculations. Figure 2 gives a snapshot of the shock diffraction patterns in water that are generated from the interaction of a Mach 1.4 shock wave with an air cavity. In Figs 3 and 4, the deformation of a copper rod and a hemispherical groove (in copper),



*Fig. 1. Mach 1.3 shock impacting a rigid wedge.*



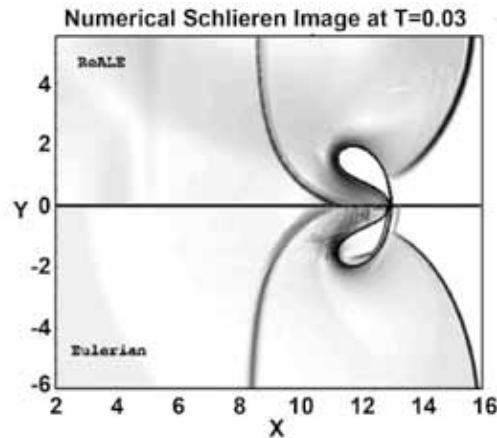


Fig. 2. Shock diffraction patterns in water.

due to high-impact conditions, are presented. The latter example is a prototype for analyzing explosively formed projectiles. Throughout these and other computations that we have investigated, the new ReALE formulation is found to be at par with existing Eulerian methods.

The development of this new paradigm, which dramatically reduces physical, numerical, and mathematical anomalies, is an important research issue that will affect multiple DOE programs and

codes. This effort will provide the necessary research on algorithms for treating hypo/hyperelastic plastic materials and reconstruction and remapping of the tensor quantities required to advance new and existing hydrocodes. Specifically, the research and resulting algorithms will be of immediate use in Eulerian, Lagrangian, and ALE hydrocodes developed under the Advanced Simulation and Computing Program. In the future, we will be extending the formulation to accurately treat embedded material interfaces. Furthermore, efforts are already underway for solving 3D applications using the advanced computing facilities available at LANL.

Fig. 4. Formation of jet due to the impact of shock wave on a hemispherical groove in copper.

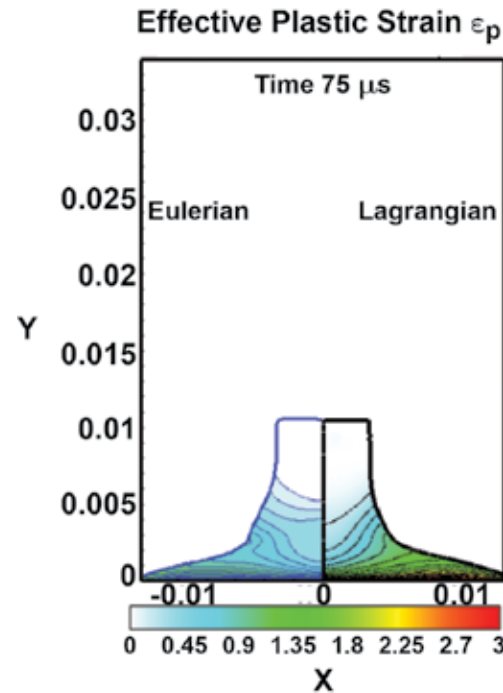
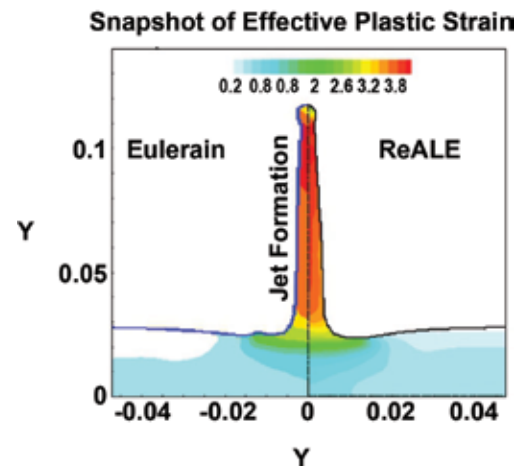


Fig. 3. Deformation of copper rod impacting rigid wall at 400 m/s.



[1] Loubere, R. et al., *J Comput Phys* **229**(12), 4724 (2010).

[2] Sambasivan, S. et al., "A Cell-Centered Lagrangian Finite Volume Approach for Computing Elasto-Plastic Response of Solids in Cylindrical Axisymmetric Geometries," LA-UR 12-20947, *J Comput Phys*, **237**, (2013).

[3] Sambasivan, S. et al., "Simulation of Collapse and Fragmentation Phenomena in a Sharp Interface Eulerian Setting," LA-UR 12-23179, *J Comput Phys*, **235**, 334 (2013).

# Minimal Stencil Finite Volume Scheme with the Discrete Maximum Principle

Konstantin Lipnikov,  
Daniil Svyatskiy, T-5;  
Yuri Vassilevki, Russian  
Academy of Sciences (RAS)

The maximum (or minimum) principle is an important property of linear and nonlinear advection-diffusion type problems. It is very desirable to mimic this property in numerical simulations in a wide range of applications. Violation of the discrete maximum principle (DMP) leads to non-physical solutions with numerical artifacts, such as a heat flow from a cold material to a hot one. These oscillations can be significantly amplified by the non-linearity of the physics. Unfortunately, numerical schemes that satisfy the DMP impose severe limitations on mesh geometry and problem coefficients. We developed a new nonlinear finite volume method that guarantees the DMP for numerical solutions on general polygonal meshes for diffusion-advection problems with anisotropic coefficients.

The diffusion-type problem can be written in the divergence form as follows:

$$\operatorname{div} \mathbf{q} = f \quad \text{and} \quad \mathbf{q} = -\mathbb{K} \nabla c + v c$$

where  $\mathbb{K}(\mathbf{x})$  is a symmetric positive definite continuous (possibly anisotropic) diffusion tensor,  $v$  is a velocity,  $f$  is a source term and  $c(x)$  is an unknown scalar function, for example concentration. The vector function  $\mathbf{q}(\mathbf{x})$  is an unknown concentration flux. The first equation represents the mass conservation law. The second is the constitutive equation that establishes connection between scalar and vector unknowns. In subsurface modeling, this law is referred to as Darcy's law—similar laws go by different names in other areas. An appropriate boundary condition should be imposed to make the problem well posed.

State-of-art second order discretization schemes, such as the Mixed Finite Element (MFE) method, Mimetic Finite Difference (MFD) method, and Multi-Point Flux Approximation (MPFA) method, fail to provide a solution that satisfies the discrete maximum principle or even to preserve positivity of a numerical solution when the diffusion tensor is heterogeneous and anisotropic or the computational mesh is strongly perturbed.

The finite volume framework has several obvious advantages. It operates with cell-centered degrees of freedom (minimum number of unknowns) and provides the local mass conservation by construction because the equation is discretized in the mixed form. The classical *linear* two-point (FV) scheme defines a flux across a mesh edge as the difference between two concentrations at neighboring cells multiplied by a transmissibility coefficient. It results in a linear system of equations with a matrix that has special properties (M-matrix with diagonal dominance in rows). This immediately implies the discrete maximum principle. These properties along with the minimal discretization stencil (number of non-zero entries

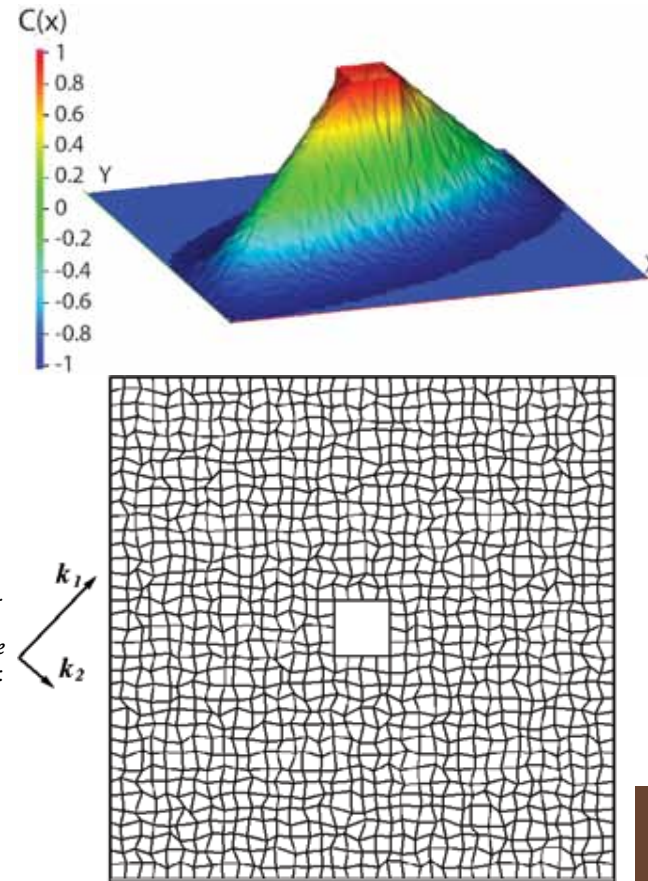
in each matrix row) make this approach very popular in many modeling tools and legacy codes. However, the accuracy of this scheme depends on the mesh geometry, mutual orientation of the mesh edges, and principal directions of the diffusion tensor. More precisely, to provide minimal order of accuracy the principal directions have to be orthogonal to the mesh edges, which is clearly an impossible requirement for arbitrary tensors and/or arbitrary polygonal cells. The MPFA scheme solves the accuracy problem by using more than two points in the flux stencil and a matrix of transmissibility coefficients. This makes the discretization stencil larger. The MPFA scheme provides a second-order accurate approximation, but is often only conditionally stable and conditionally monotone.

To incorporate the monotonicity requirement into the finite volume framework, we use the ideas proposed by E. Bertolazzi, [1] along with the fact that coefficients in flux discretization depend on the unknowns in neighboring cells even for linear diffusion-advection problems. Several approaches based on this idea have been proposed recently, but all of them guarantee only positivity preservation of a numerical solution on general meshes for general tensor coefficients. To guarantee the DMP, a *nonlinear* multi-point approximation of the flux is essential. For diffusion problems, a new method was proposed that uses multi-point approximation of the flux along with interpolation techniques and auxiliary unknowns at the mesh vertices [3]. The use of interpolation techniques and auxiliary unknowns increases the stencil and makes it difficult to incorporate this approach into existing modeling tools. In our research, we propose an interpolation-free multi-point nonlinear approximation of diffusive fluxes. The proposed scheme has the minimal stencil and reduces to the classical two-point FV scheme on Voronoi or rectangular meshes for scalar (and, in a few cases, diagonal tensor) coefficients.

The resulting nonlinear algebraic system is sparse, but non-symmetric. For quadrilateral meshes there are at most four non-zero elements in each row. We elaborated on this system and proved, theoretically as well as demonstrating in the numerical tests (see Fig. 1), that the numerical solution satisfies the DMP principle. Moreover, if the nonlinear system is solved using the Picard iterative method then the proposed method guarantees that the DMP is satisfied on all iterative steps. It means that for any tolerance of the iterative method we obtain a monotone solution. The proof is based on the special properties of M-matrices.

The numerical experiments presented in [2] also study this approach for advection-diffusion problems and demonstrate its monotone properties and accuracy. The method can be applied on unstructured polygonal meshes and full anisotropic heterogeneous diffusion tensors. The second-order convergence is observed for scalar unknowns  $C_h$ . For numerical approximation  $q_h$  of the normal components of the flux  $q$  the convergence rate, is higher than the first order.

Fig. 1. (Top): Profile of numerical solution  $C_h(x,y)$  on the distorted quadrilateral mesh,  $-1 \leq C_h(x,y) \leq 1$ . (Bottom): Computational domain for the anisotropic diffusion problem: The unit square with the hole in the center. The problem becomes the diffusion equation with highly anisotropic tensor. Ratio of tensor's eigenvalues is  $10^3$ . Tensor is rotated with respect to coordinate axes on  $60^\circ$  clockwise.  $C_h=1$  on the hole,  $C_h=-1$  on the boundary of unit square. Analytical solution satisfies maximum principle,  $-1 \leq c(x,y) \leq 1$ .

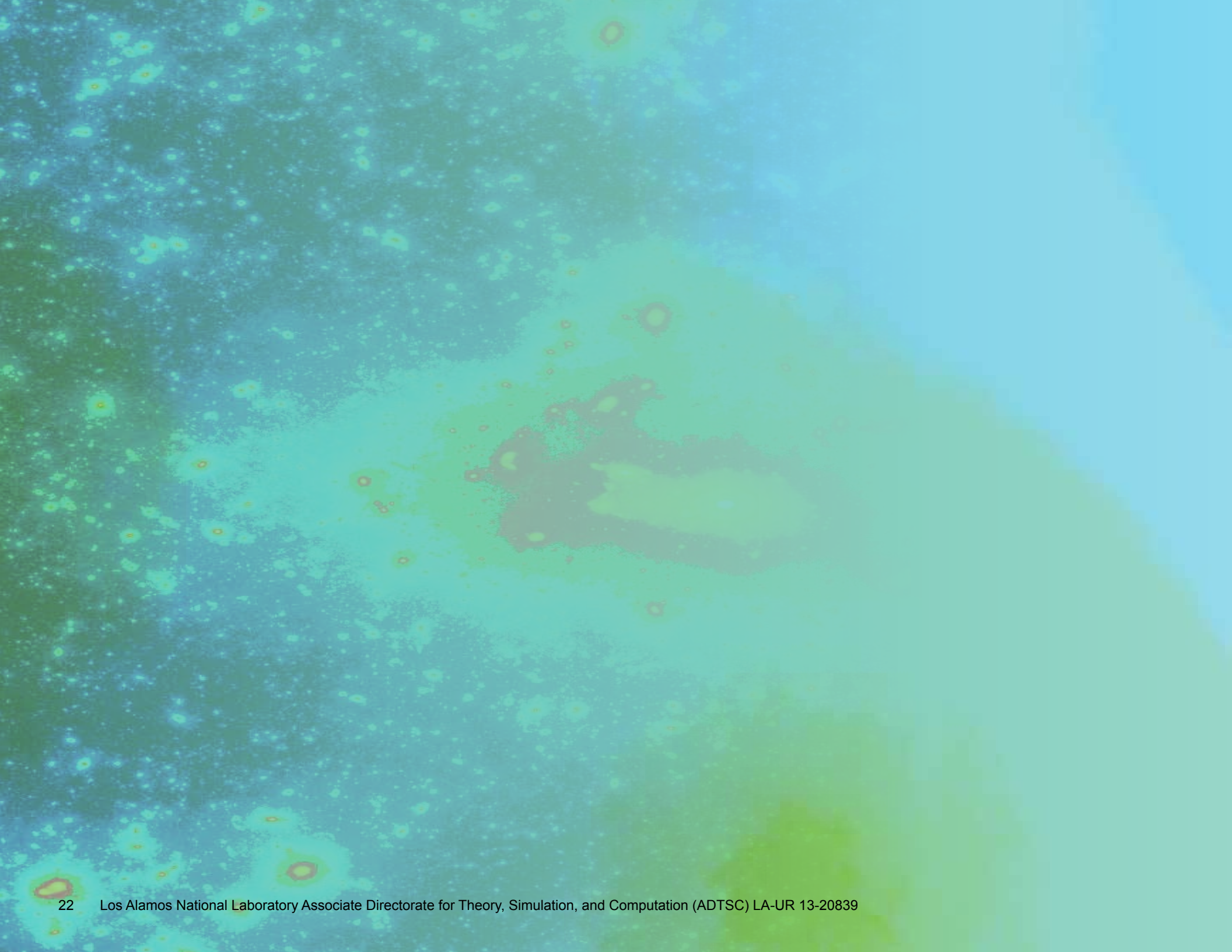


[1] Bertolazzi, E. *Math. Mod. Meth. Appl. S.*, **8**,(4), 685 (1998).

[2] Sheng, Z. and G. Yuan, *J Comput Phys* **230**(7), 2588 (2011).

[3] Lipnikov, K. et al., *Russ. J. Numer. Anal. Math. Modelling*, **27**(4), 369 (2012).





# Astrophysics and Cosmology

The work described in this section involves collaborations throughout the Theory, Simulation, and Computation Directorate, across LANL, and worldwide. These articles represent only a small set of examples of the great breadth and broad interest in Astrophysics and Cosmology at LANL. Articles in this section cover subjects focused on the most exotic objects in the universe, including supernova progenitors, supernova remnants, the first planets in the universe, supermassive stars, and cosmological simulations of large-scale structure in the universe. These research efforts are excellent examples of the high-quality basic research conducted in the Directorate and indications of the overlap with programmatic activities of critical importance to fulfilling LANL's nuclear security mission.

# Studying Supernova Progenitors with Observations and Simulations

Lucille H. Frey, HPC-3,  
University of New Mexico;  
Wesley Even, XTD-6;  
Amanda J. Bayless,  
Southwest Research Institute;  
Chris L. Fryer, CCS-2,  
University of Arizona,  
University of New Mexico;  
Peter W.A. Roming,  
Southwest Research  
Institute, Pennsylvania State  
University;  
Tyler Pritchard, Pennsylvania  
State University

With increasing numbers of supernovae (SNe) being discovered by survey telescopes, the amount of observational data is rapidly increasing. Most of this data, lightcurves and spectra, provide only indirect information about the progenitor stars and their environments. Comparing simulations to these observations allows us to study the interior composition and other difficult-to-observe properties of the supernova (SN) progenitors. We have used this method to study the recently discovered SN 2012aw and are comparing observations of SN 1987A to our simulations to determine the accuracy of our technique. We are creating a database of simulations that will allow us to more easily match observed supernovae to simulations and derive progenitor and explosion properties.

Supernovae are being observed in increasing numbers, with hundreds discovered each year by a wide array of survey telescopes [1]. Despite the large amounts of data, the information we can derive about the progenitor stars from all of these observations is limited because for most of these SNe we have only lightcurve data, while for some we have spectra as well. Even in the extremely rare cases where the progenitor stars of nearby SNe can be identified on pre-explosion images[2], this only gives a direct measurement of the luminosity and color of the star. The photospheric radius, mass, and temperature of the star can then be inferred—these observations do not give any detail about the interior composition of the star or, except in very extreme cases, its mass loss history. By comparing self-consistent simulations to observed SNe, we can learn more about the composition and size of the progenitor, its mass loss history, and characteristics of the SN explosion itself.

One recent SN, 2012aw, was near and bright enough to allow ultraviolet (UV) and optical observations to be made over several months [3,4]. Previous observations of the same region give a constraint of  $\pm 0.8$  days for the initial explosion time. This means that the explosion time is not a free parameter, as it is for many SNe when the rising phase is not observed and pre-explosion observations do not exist. SN 2012aw was classified as a Type II-P because of the flat plateau observed in the optical lightcurve. It has a UV lightcurve that appears similar to other SNe at early times, but with a flatter slope after about a month (see Fig. 1).

Several groups have used pre-explosion images to obtain estimates of the

progenitor's properties, giving best estimates of the mass and radius of  $15\text{--}20\text{ Msun}$  and  $1040 \pm 100\text{ Rsun}$  [5,6].

We have developed a pipeline to simulate supernovae, beginning with stellar evolution models, simulating core-collapse and then using the rad-hydro code RAGE [7] to follow the evolution of the shock as it propagates outward. We can then calculate spectra at any point after core collapse and integrate over those spectra to create lightcurves for any wavelength band [8].

To learn more about 2012aw and to test the physical accuracy of our new simulation pipeline, we ran a series of simulations and compared their lightcurves to 2012aw. We began with a previously run 2-Msun progenitor that showed similar lightcurve characteristics and modified the wind profile to obtain a closer match. We also tried a smaller 15-Msun progenitor that was intended to be a II-P progenitor, but has a lightcurve that drops off much earlier than that observed for 2012aw. The best fit we have obtained so far is the 23-Msun progenitor, with a mass loss rate of  $1e-5\text{ Msun/year}$  and a wind velocity of  $1e7\text{ cm/s}$  (see Figs. 2 and 3). The radius of this progenitor is  $650\text{ Rsun}$ , smaller than the estimated value for the 2012aw progenitor, so we may obtain a better fit starting with a larger stellar radius.

While comparing our simulations to observed SNe can tell us about the SN being observed, the number of variables involved in the simulations means that it would be easy to obtain similar results with multiple combinations of progenitor properties. To perform a more controlled test of the accuracy of our code, we are beginning a project to match a simulation to SN 1987A. We chose this SN as a test case since it has been very consistently and thoroughly observed since it first exploded, due to its nearby location in a satellite galaxy of the Milky Way, the Large Magellanic Cloud. We will be able to create our progenitor star and the wind environment using known or well-constrained parameters

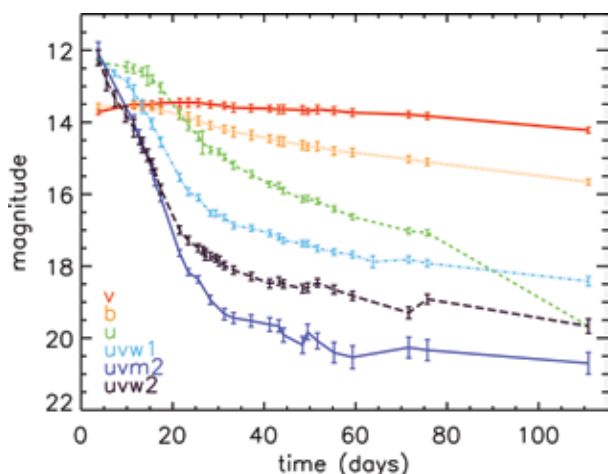


Fig. 1. Observed UV/Optical lightcurve for 2012aw.



and will have many fewer unknowns to adjust. The large amounts of observational data and previously performed analysis will also provide a very comprehensive set for comparisons for both lightcurves and spectra.

The models run for these comparisons to SN2012aw and SN1987A are being used to start a database of SN simulations that will eventually allow future observations to be matched to simulations much more easily. This database will eventually include SN simulations with various combinations of SN type, progenitor mass, metallicity, wind profile, explosion energy, and internal stellar composition (due to different mixing algorithms).

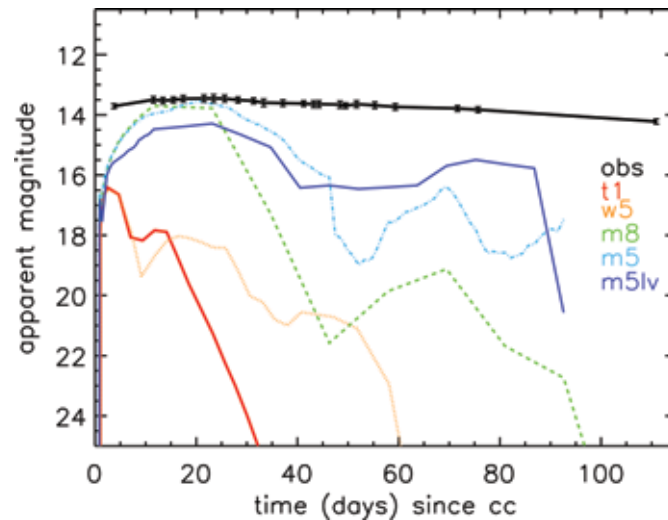


Fig. 2. V-band lightcurve for 2012aw, compared to simulated lightcurves.

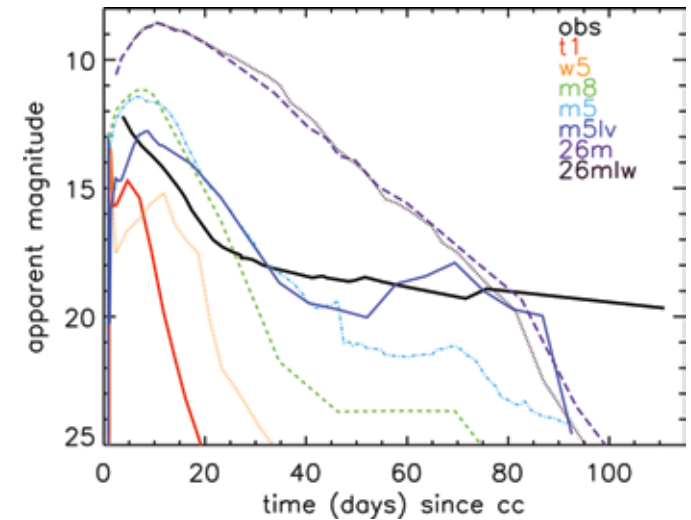


Fig. 3. uvw1-band lightcurve for 2012aw, compared to several simulated lightcurves.

- [1] IAU Central Bureau for Astronomical Telegrams, List of Supernovae, <http://www.cbat.eps.harvard.edu/lists/Supernovae.html>
- [2] Smartt, S.J., *Annu Rev Astrpm Astrophys* **47**, 63 (2009).
- [3] Fagotti, P., *CBET* **3054**, 1 (2012).
- [4] Bayless, A.J. et al., arXiv:1210.5496. (2013).
- [5] Van Dyk, S.D. et al., *Astrophys J* **756**, 2 (2012).
- [6] Fraser, M. et al., *Astrophys J Letters* **759**, 1 (2012).
- [7] Gittings, M. et al., *CS&D* **1**, 015005 (2008).
- [8] Frey, L.H. et al., arXiv:1203.5832. (2013).

# Supernova Remnants: Nature's Turbulence Experiment

Chris L. Fryer, CCS-2;  
Carola Ellinger, University of  
Texas;  
Andreas Zoglauer, University  
of California

A supernova remnant is the term used for the brilliant display produced when the debris of a supernova explosion strikes its immediate surroundings. These surroundings are determined by the last stages of evolution of the massive star progenitor of the supernova explosion. As such, supernova remnants not only provide a unique window into our understanding of supernovae but also of stellar evolution and turbulent mixing.

Scientists at LANL have formed a multi-institutional team to study the evolution of supernova remnants, focusing on the nucleosynthetic yields and the explosive mixing as these nuclear products interact turbulently with the surrounding stellar winds.

This year, Carola Ellinger, working with LANL scientists Chris Fryer and Gabriel Rockefeller, conducted a detailed study of the mixing in asymmetric supernovae [1]. Figure 1 shows maps of nuclear abundances at the end of the mixing phase caused as the shock moves through the star. This mixing explains many of the observed features in supernovae

like SN1987A where both the hydrogen and helium showed evidence of mixing inward and the iron was mixed outward.

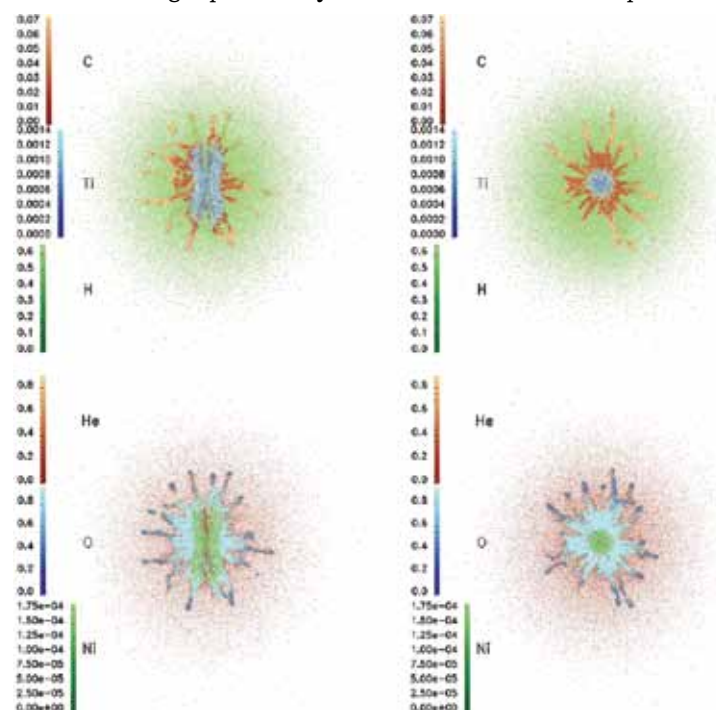
One of the most striking supernova remnants is the roughly 330-year-old Cassiopeia A remnant. In 2004, LANL was part of a team that observed Cassiopeia A for a megasecond image on the Chandra satellite [2] (a full 10 days focused on this remnant). These observations provide detailed information of the remnant characteristics, producing spectral data at each pixel of a wide map. Figure 2 shows the image from these observations. First note the detailed turbulent structures produced as the shock plows through the interstellar medium. With the spectral data, scientists are able to determine both the density and temperature of all features in this image. No other experiment, laboratory or astrophysical, can provide such detailed diagnostics of a turbulent event.

The difficulty with using this data lies in understanding (or not understanding) the initial conditions. These initial conditions require both an understanding of the stellar progenitor and the subsequent supernova explosion. LANL was part of the leading effort to model the progenitor of Cassiopeia A [3], placing strong constraints on the mass of the progenitor. From this work, it was realized that the progenitor of Cassiopeia A was definitely a binary system. This makes modeling a bit more difficult, because asymmetries in the remnant profile

(e.g., the apparent “jet” in the Cassiopeia A remnant) could be due to the asymmetries in the circumstellar material, not in asymmetries in the explosion itself.

One way to distinguish circumstellar density asymmetries from explosion asymmetries is to study the asymmetries in material produced in explosive nucleosynthesis, such as  $^{56}\text{Ni}$  or  $^{44}\text{Ti}$ . NuSTAR (the Nuclear Spectroscopic Telescope Array) is the first focusing telescope in the high energy X-ray band (6–79 keV) of the electromagnetic spectrum.

*Fig. 1. Abundance maps for a jet-like explosion where the velocity along the axis is four times greater than in the equator. The figure shows slices that are parallel and perpendicular to the polar axis. Hydrogen and helium are visibly mixed inward along the asymmetry axis, while nickel (which will decay to iron) and titanium are mixed somewhat closer into the Rayleigh-Taylor fingers.*



The decay chain of radioactive  $^{44}\text{Ti}$  produces photons in this band, and NuSTAR will be able to map out the  $^{44}\text{Ti}$  distribution in supernova remnants. LANL has joined the science team of NuSTAR to produce  $^{44}\text{Ti}$  distributions for supernova remnants. Figure 3 shows our first calculation for Cassiopeia A, modeled through the NuSTAR sensitivity band [4]. NuSTAR has now spent over 500 ks observing Cassiopeia A and we expect to finish, within the next six months, a detailed comparison of this NuSTAR data to our simulations.

With the NuSTAR data, we will have firm constraints on the supernova explosion asymmetries. Combined with the data from a host of other satellites, we will finally have constrained initial conditions for Cassiopeia A. With these constraints, we will be able to extract the wealth of physical information from this cosmic physics experiment.

Fig. 2. Three-color image of Cassiopeia A with red corresponding to the Silicon  $\text{He}\alpha$  (1.78–2.0 keV) line, blue corresponding to the Fe K (6.52–6.95 keV) line, and green corresponding to the 4.2–6.4 keV continuum. (Bottom left): Overexposed broadband image showing faint features. The spectral regions are indicated (Top left): northeast jet; (Bottom right): on the same scale, the ratio image of the Si  $\text{He}\alpha$  (1.78–2.0 keV) and 1.3–1.6 keV (Mg  $\text{He}\alpha$ , Fe L), without subtraction of the continuum contribution. The image highlights the jet and counterjet traced by Si emission, although features at the lowest intensity levels are uncertain.

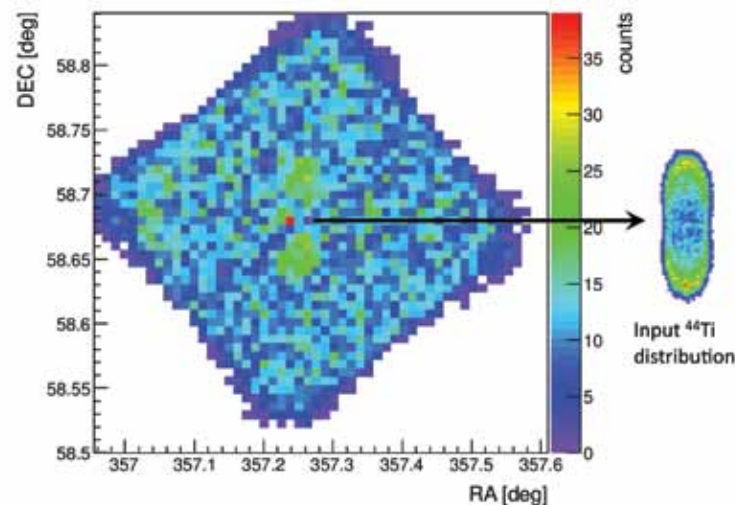
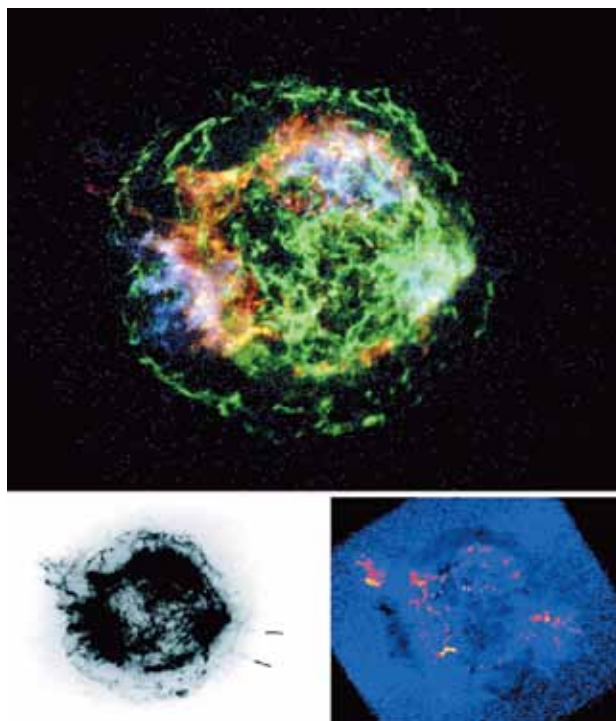


Fig. 3. Color map (denoting intensity) of an asymmetric explosion of Cassiopeia A as viewed by NuSTAR. The asymmetric explosion is a jet-like explosion like that shown in Fig. 1. (The velocity is four times stronger along the axis than the equator.)

- [1] Ellinger, C.I. et al., *Astrophys J* **755**, 160 (2012).
- [2] Hwang, et al., *Astrophys J* **615**, L117 (2004).
- [3] Young, et al., *Astrophys J* **640**, 891 (2006).
- [4] Zoglauer et al., American Astronomical Society, HEAD meeting #12, #43.07 (2012).



# The First Planets

Jarrett L. Johnson,  
Hui Li, T-2

We explore the conditions under which, according to the two leading theories of planet formation, the first planets formed in the early universe. In the context of the core accretion model for planet formation, we estimate that the minimum abundance of heavy elements required for planet formation is  $[\text{Fe}/\text{H}]_{\text{crit}} \sim -1.5 + \log(r/1 \text{ AU})$ , where  $r$  is the distance between the planet and its host star, an astronomical unit (AU) is the distance between the Earth and the Sun, and here the iron abundance relative to that of the Sun  $[\text{Fe}/\text{H}]$  is a proxy for the abundance of heavy elements. This minimum value implies that the first Earth-like planets likely formed around stars with heavy element abundances at least one-tenth that of the Sun. We find, however, that the first gas giant planets may have formed via gravitational instability in circumstellar disks, instead of via core accretion, although in this case they must have formed within a few AU of their host stars. We show that the available observational data are consistent with these predictions.

Following the formation of the first stars and galaxies, the formation of the first planets marks a milestone in the increasing complexity of the cosmos. When and where the first planets form depends in large part on the mechanism by which they form—there are two main models for how planet formation takes place. In the core accretion model, dust grains coagulate and settle into the mid-plane of a circumstellar disk, where they merge to form planetesimals and eventually the cores of planets. This is the prevailing model for how rocky, terrestrial planets take shape, although if the cores that form accrete large amounts of gas then gas giant planets may form as well [1]. In the gravitational instability model, a circumstellar disk grows to the point that it fragments into planets under its own gravity—it is thought that this model can explain the formation of some gas giant planets, although it cannot account for terrestrial planets [2].

When the abundance of heavy elements is very low in a circumstellar disk, as is the case in the early universe when only the earliest supernovae have chemically enriched the gas, there are two factors that work against planet formation in the core accretion model. First, it takes much longer for dust grains, of which there is a low abundance, to collide with one another, grow, and settle into the mid-plane of the circumstellar disk [3]. Second, due to the lower opacity to ultraviolet (UV) and X-ray photons emitted from the host star, the circumstellar disk is photoevaporated away much more quickly than its more chemically enriched counterparts, and once the disk is destroyed planet formation can no longer take place [4]. Therefore, there is a competition between these two time scales, and it is only above some minimum, or “critical,” heavy element abundance that there is time for dust grains to coagulate

and form planetesimals before the circumstellar disk is destroyed by radiation from the host star [5]. Figure 1 shows this critical heavy element abundance, expressed as the iron abundance relative to that of the Sun  $[\text{Fe}/\text{H}]$ , as a function of distance  $r$  from the host star. Below the critical abundance planet formation is not possible and so this is termed the “forbidden zone” for planet formation via core accretion. The data on several hundred observed planets [6] are also shown, all of which lie safely outside the forbidden zone, indicating that they may have formed via core accretion.

In the other main model of planet formation, gravitational instability, it is possible that gas giant planets could form in circumstellar disks with heavy element abundances below the critical value for the core accretion model. Indeed, the recently reported detection of a gas giant planet on a wide orbit around a heavy element-depleted star provides some indication that this may take place [7]. While planet formation via gravitational instability is not likely to occur in the disks surrounding the first stars, as they are much too hot to be subject to gravitational fragmentation at the small scale of planets, slightly chemically enriched circumstellar disks could cool sufficiently for this mode of planet formation to function [8]. In the early universe, however, there is a fundamental lower limit to the temperature to which disks can cool radiatively that is set by the cosmic microwave background (CMB) radiation. Because the temperature of a circumstellar disk cannot go below the temperature of the CMB, planets could only form via gravitational fragmentation relatively close to their host stars—outside some maximum radius there is a second forbidden zone within which planet formation via gravitational instability is not possible. Because the temperature floor set by the CMB is higher

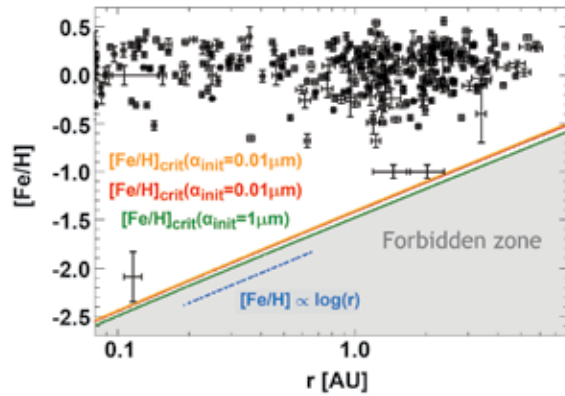


Fig. 1. The critical metallicity for planet formation, expressed as the iron abundance relative to that of the Sun  $[Fe/H]$ , as a function of distance  $r$  from the host star. The three colored lines correspond to different initial dust grain sizes, as labeled. The curves shown here are well approximated by  $[Fe/H]_{crit} \sim -1.5 + \log(r/1 \text{ AU})$ , where an AU is the distance between the Earth and the Sun. The black crosses show the iron abundance  $[Fe/H]$  of observed planet-hosting stars plotted against the semi-major axes of the planets' orbits. Planetary systems that lie below this line, in the forbidden zone shown in gray, are unlikely to have formed through core accretion, at least not at their present locations.

at earlier cosmic times, this maximum radius is smaller at earlier epochs, as shown in Fig. 2. Also shown are the data on gas giant planets orbiting stars with low heavy element abundances. The vast majority of the planets lie well below the forbidden zone, although there are a few very old planets and planets on very wide orbits that lie outside of it. Thus, most of the planets shown could have formed via gravitational

instability unhindered by the constraints placed by the CMB temperature floor—those few in the forbidden zone, however, may have formed via core accretion instead.

In particular, the single heavy element-depleted gas giant planet recently reported to lie in the forbidden zone for core accretion [7] lies well outside of the forbidden zone for gravitational instability. This suggests that the first planets may have been gas giants formed via gravitational instability, although the first terrestrial planets must have instead formed via core accretion. In turn, the first terrestrial planets as large as our Earth, which may have hosted the first life in the cosmos, are expected to have formed via core accretion only once the heavy element abundance of circumstellar disks was at least a tenth of that of the Sun [5].

#### Special Thanks

We thank Volker Bromm, Dave Collins, Pawan Kumar, Doug Lin, Alexia Schulz, and Tsing-Wai Wong for helpful discussions.

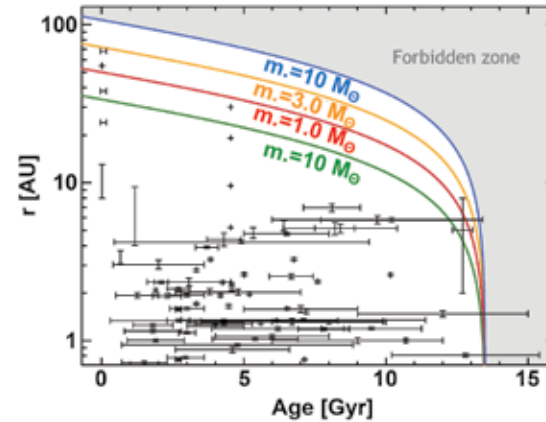


Fig. 2. The semi-major axes (vertical axis) of the orbits, and host stellar age (horizontal axis), of observed planets are shown as black crosses. The colored lines show the maximum possible distance  $r_{max}$  at which planets can form from their host stars via gravitational instability as a function of their present age, for four different host stellar masses, as labeled. Beyond this maximum distance, it is predicted that planet formation is not possible via gravitational instability due to the temperature floor set by the CMB, and so this region of the plot is termed the forbidden zone for planet formation via gravitational instability.

- [1] Pollack, J.B. et al., *Icar* **124**, 62 (1996).
- [2] Boss, A.P., *Science* **276**, 1836 (1997).
- [3] Dullemond, C.P., and C. Dominik, *Astron Astrophys* **434**, 971 (2005).
- [4] Ercolano, B., and C.J. Clarke, *Mon Not R Astron Soc* **402**, 2735 (2010).
- [5] Johnson, J.L. and H. Li, *Astrophys J* **751**, 81 (2012).
- [6] Wright, J.T., *Publ Astron Soc Pac* **123**, 412 (2011).
- [7] Setiawan, J. et al., *Astron Astrophys* **540**, 141 (2012).
- [8] Johnson, J.L. and H. Li, Arxiv e-prints, 1212.1482 (2012).

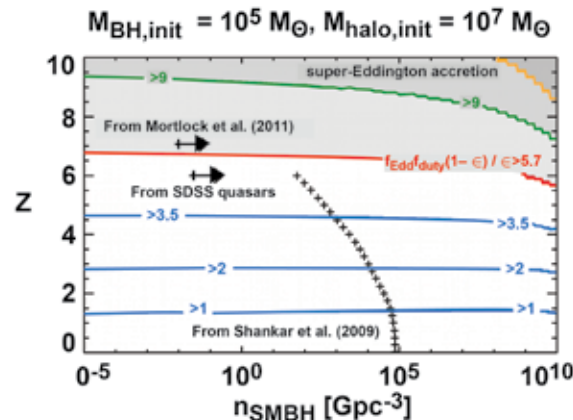
# A Strong Case for Supermassive Stars in the Early Universe

Jarrett L. Johnson, T-2;  
Daniel J. Whalen, Carnegie  
Mellon University;  
Hui Li, T-2;  
Claudio Dalla Vecchia,  
Bhaskar Agarwal,  
Sadegh Khochfar,  
Max Planck Institute for  
Extraterrestrial Physics;  
Daniel E. Holz, University of  
Chicago

Since the detection of supermassive black holes (SMBH) a decade ago, it has been a challenge to explain how they have reached masses greater than  $10^9$  solar masses ( $M_\odot$ ) less than a billion years after the Big Bang. Drawing on a growing body of theoretical and observational evidence, we argue that the “seed” black holes (BH) from which they grew must have had masses of at least  $\sim 10^5$   $M_\odot$ , consistent with their formation by direct collapse in primitive galaxies. Corroborating this, some of the most realistic large-scale cosmological simulations to date show that massive SMBH seeds could have formed much more often in the early universe than previously assumed. We conclude that these may have been the origin of most SMBHs in the centers of galaxies today and that they may soon be detected in upcoming deep surveys by the James Webb Space Telescope (JWST).

Fig. 1. The rate and radiative efficiency of accretion onto various BH seeds needed to produce SMBHs with masses  $>10^9 M_\odot$  with a space density  $>n_{\text{SMBH}}$  (horizontal axes) by redshift  $z$  (vertical axes) for two different scenarios for the initial BH seeds:  $10^5 M_\odot$  BH formed in  $10^7 M_\odot$  dark matter halos (top) and  $10^2 M_\odot$  Pop III progenitors formed at earlier times in  $10^5 M_\odot$  halos (bottom). The yellow, green and red contours correspond to radiative efficiencies  $\epsilon = 0.07, 0.1$ , and  $0.15$ , for the case of constant accretion at the Eddington rate; the three shaded regions show the cases in which super-Eddington growth is required, for these three radiative efficiencies. Also shown are the space densities of SMBHs with masses  $>10^9 M_\odot$  inferred from observations. These data are most easily explained if the seeds of the highest-redshift SMBHs formed from SMSs, instead of from Pop III seeds. This provides indirect evidence for the existence of SMSs in the early universe.

Over the past decade, observations have revealed that quasars powered by the accretion of gas onto SMBHs are assembled at very early times in cosmic history. The most distant quasar known, at a redshift  $z \sim 7$ , has been recently inferred to harbor a SMBH with a mass of  $\sim 2 \times 10^9 M_\odot$  [1]. How such a BH grew to such masses less than 800 million years after the Big Bang remains a mystery. There are two main competing theories for the origin of SMBHs [2]. The first posits that an initial  $\sim 100 M_\odot$  “seed” BH formed from the collapse of primordial (or Pop III) stars that formed in primordial gas  $\sim 200$  million years after the Big Bang. The second posits that a much more massive,  $\sim 10^5 M_\odot$  seed BH formed during the catastrophic collapse of gas at the centers of primitive galaxies  $\sim 500$  million years after the Big Bang. Once formed, these seeds are thought to rapidly accrete gas and grow into the SMBHs that power the  $z \sim 7$  quasars.



In both of these models, the BHs must accrete gas at or above the Eddington limit (the accretion rate at which pressure from the radiation in the accretion disk around the BH balances gravity) in order to grow quickly enough to explain the highest-redshift quasars. Not only this, but the radiative efficiency of the BH, defined as the fraction of the rest mass energy of accreted material that is converted to radiation in the accretion disk, must be low as well. Figure 1 shows the radiative efficiencies  $\epsilon$  of the observed population of SMBHs with masses  $>10^9 M_\odot$  that permit the formation of these behemoths in the two dominant BH seed theories (top and bottom panels). For Pop III seeds (bottom panel) Fig. 1 shows that the maximum allowed  $\epsilon$  is  $\sim 0.07$ , while for massive SMBH seeds (top panel) the maximum  $\epsilon$  is higher,  $\sim 0.1$ . These limits strongly favor the massive BH model as the most viable because  $\epsilon$  for high-redshift SMBHs is generally inferred to be  $>0.1$  [3]. In some cases,  $\epsilon$  is thought to be as high as  $0.3$ – $0.4$ , which would imply accretion at rates exceeding the Eddington limit. In any case, however, the observational constraints are most easily satisfied by massive SMBH seeds because they can grow more quickly than less massive BHs.

Independent evidence for the formation of massive SMBH seeds by direct collapse comes from recent numerical simulations by independent research groups that suggest that these objects formed more frequently than previously assumed. It is generally thought that catastrophic gas collapse can only occur in protogalaxies immersed in very high molecule-dissociating UV backgrounds that suppress  $H_2$  cooling until far more rapid atomic cooling begins [4,5]. It has been assumed in the past that these conditions are rarely met, since the stellar populations that would produce the large UV fluxes would also likely give rise to supernovae that would chemically enrich the primordial gas. However, recently the



flux that is required to keep the protogalaxies molecule-free has been shown to be lower than previously expected [6,7]. In turn, as shown in Fig. 2, patches of the universe where primordial gas is subjected to such high UV fields have been found in the first large-scale cosmological simulations that model chemical enrichment by supernovae and the build-up of the molecule-dissociating radiation field self-consistently [8].

Semi-analytical models likewise suggest that massive SMBH seeds readily formed in the early universe [9,10]. As shown in Fig. 3, if the progenitors of these seeds were supermassive stars (SMS), there would be enough of them to be detected at redshifts  $z > 6$  in deep-field surveys by the JWST for a variety of molecule-dissociating UV backgrounds. Indeed, enough SMSs may have formed at high redshifts to be the seeds of most of the SMBHs inferred to inhabit the centers of galaxies today [10].

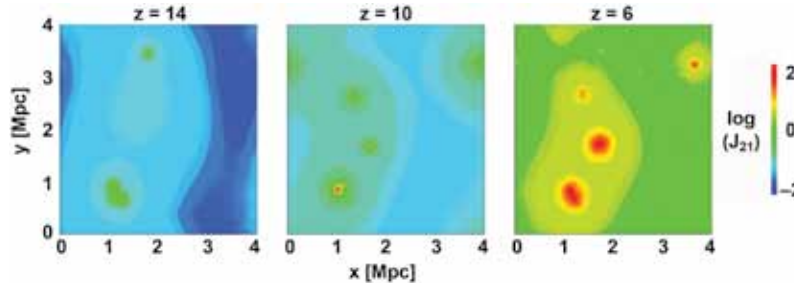


Fig. 2. The flux of molecule-dissociating radiation in units of  $J_{21}$  ( $10^{21} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ Hz}^{-1} \text{ sr}^{-1}$ ) generated by the stars formed in a  $(4 \text{ Mpc comoving})^3$  cosmological simulation volume, at three redshifts:  $z = 14$  (left), 10 (middle), and 6 (right). Massive SMBH seeds can form in regions that are not yet chemically enriched by supernova explosions and which are exposed to fluxes  $J_{21} > 30$ . Such regions are found to exist in recent cosmological simulations.

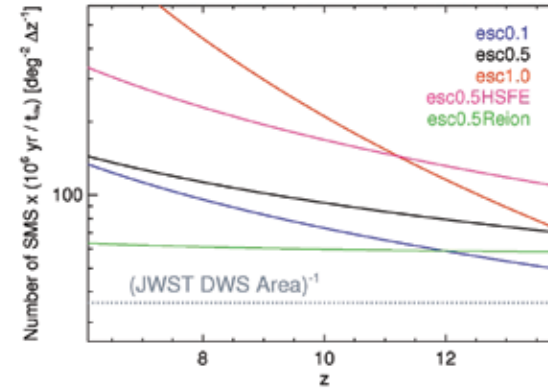


Fig. 3. The number of SMSs, as observed on the sky per square degree per redshift interval ( $\Delta z$ ), as a function of redshift  $z$ . The number of SMSs is shown for five simulations in which different prescriptions have been used for star formation and radiative feedback. The dotted gray line shows the number of SMSs that must be present for at least one per redshift interval ( $\Delta z = 1$ ) to appear in the field of view of the Deep-Wide Survey planned for the JWST. For all cases, the survey should be large enough to detect at least a few SMSs.

### Special Thanks

Thanks to Xiaohui Fan, Chris Fryer, and Marta Volonteri for valuable feedback on the work reported here.

- [1] Mortlock, D.J. et al., *Nature* **474**, 616 (2011).
- [2] Volonteri, M., *Science* **337**, 544 (2012).
- [3] Johnson, J.L. et al., Arxiv e-prints, 1211.0548 (2012)
- [4] Hosokawa, T. et al., *Astrophys J* **756**, 93 (2012).
- [5] Johnson, J.L. et al., *Astrophys J* **750**, 66 (2012).
- [6] Shang, C. et al., *Mon Not R Astron Soc* **402**, 1249 (2010).
- [7] Wolcott-Green, J. et al., *Mon Not R Astron Soc* **418**, 838 (2011).
- [8] Johnson, J.L. et al., *Mon Not R Astron Soc* **428**, 1857 (2012).
- [9] Petri, A. et al., *Mon Not R Astron Soc* **422**, 1690 (2012).
- [10] Agarwal, B., et al., *Mon Not R Astron Soc* **425**, 2854 (2012).

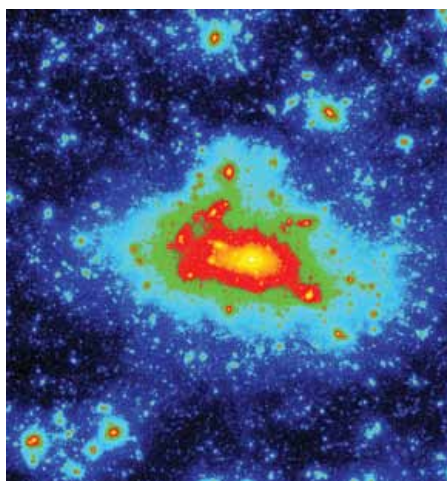
# The Hashed Oct-Tree N-Body Algorithm at a Petaflop

Michael S. Warren, T-2;  
Ben Bergen, CCS-7

Cosmological simulations are the cornerstone of theoretical analysis of large-scale structure. During the next few years, projects such as the South Pole Telescope (SPT) and the Dark Energy Survey (DES) will measure the spatial distribution of large-scale structure in enormous volumes of space across billions of years of cosmic evolution. Within a decade, the next generation of observing projects such as the Large Synoptic Survey Telescope (LSST)—in optical wavelengths—and the Square Kilometer Array (SKA)—in the radio spectrum—will gather hundreds to thousands of petabytes of observational data. Advances in modeling must keep pace with observational advances if we are to understand the universe that led to these observations.

We recently demonstrated our hashed oct-tree N-body code (HOT) scaling to 256 k processors on Jaguar at ORNL with a performance of 1.79 Petaflops (single precision) on 2 trillion particles. We also performed preliminary studies with NVIDIA Fermi graphical processing units (GPU), achieving single-GPU performance on our hexadecapole inner loop of 1 Tflop (single precision) and application performance speedup of 2× by offloading the most computationally intensive part of the code to the GPU.

*Fig. 1. A halo from a high-resolution simulation containing 134 billion particles. The image is 8 Mpc across, and the spatial resolution of 1 kpc is 1/10th the scale of a pixel in this image.*



Understanding the nature of dark matter and dark energy is undoubtedly among the most important unsolved problems in physics. The intrinsic non-linearity of the gravitational evolution of matter in the universe has limited the analytic studies of the problem to small perturbations or restricted symmetries. The only known way to obtain accurate 3D solutions is via numerical simulation. Modern N-body simulation codes provide the theoretical basis for our present understanding of the mass distribution in the universe, and are an essential link in the chain that connects particle physics to cosmology.

Such simulations have been at the forefront of parallel computing since the early 1990s. N-body simulations have grown from 300 particles in 1970 to hundreds of billions of particles today.

We have a long and distinguished history in the development of parallel numerical techniques for solving astrophysical and cosmological N-body problems. We have achieved superior performance on multiple generations of the fastest supercomputers in the world with our HOT, spanning two decades and garnering multiple Gordon Bell Prizes for significant achievement in parallel processing. With a recent Discretionary award on Jaguar at ORNL we have demonstrated our code scaling to 256 k processors with a performance of 1.79 Petaflops (single precision) on 2 trillion particles.

The revolutionary transformation of cosmology from a qualitative to a quantitative science has occurred over just the last 20 years. Driven by a powerful and diverse suite of observations, the parameters describing the large-scale universe are now known to approach 1% precision. This remarkable narrowing of parameters greatly facilitates research on the next prime target of precision cosmology: understanding the growth of cosmological structure in the non-linear regime. Our scientific aims are to use computer simulations to better understand the fundamental properties of the large-scale universe. These questions at the frontier of science include: How do cosmic structures form and evolve? What is dark matter? Why is the universe accelerating?

A great deal of progress has been made over the past 25 years in parallel computing, but much remains the same. One of our first scientific N-body simulations of dark matter in 1990 used 1.1 million particles and was performed on the Caltech/JPL Mark III hypercube in 1990. The 64-node machine used Motorola 68000 microprocessors accelerated with a Weitek floating point co-processor. The simulation was completed in 60 hours, sustaining 160 Mflops with a parallel efficiency of 85%. We have recently demonstrated our ability to use a similar algorithm to perform a simulation on 260 thousand processors with over one trillion ( $10^{12}$ ) particles, sustaining in excess of 1.5 Petaflops with a parallel efficiency of 90% (Fig. 1). Since our first parallel simulations, the message-passing programming model, time to solution, and parallel efficiency are nearly the same, but the problem size has increased by

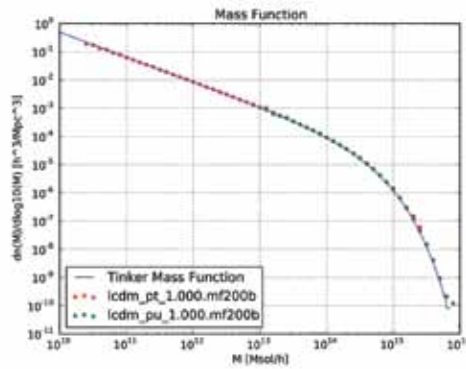


Fig. 2. The halo mass function across five orders of magnitude from three 69 billion particle simulations compared with the Tinker (2008) fit.

a factor of a million, and performance by a factor of 10 million. Increasing the performance of an automobile by a factor of 10 million would allow it to travel at the speed of light!

Our simulations are being driven to higher and higher particle numbers due to the simple fact that observations are probing smaller scales to higher accuracy, but we cannot increase our small-scale resolution at fixed particle number without reducing the spatial volume of the simulation,

which then creates errors due to the statistical variance in the large-scale modes. With a recent series of 69-billion-particle simulations performed at LANL, we can quantify these errors at both large and small scales, and calculate the simulation parameters necessary to produce data products that are accurate to 1% or better over the range of interest. An example dark matter halo from one of our high-resolution simulations is shown in Fig. 2.

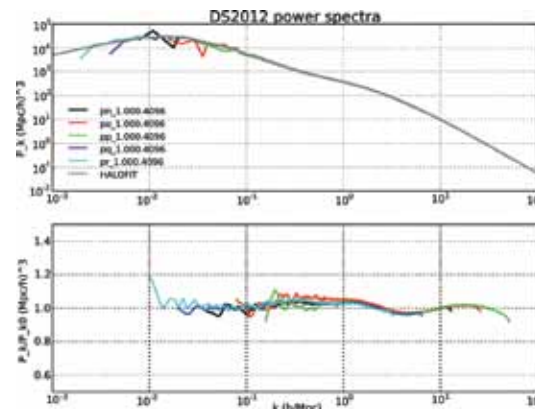


Fig. 3. The power spectrum of density fluctuations at redshift zero from six of our latest  $N=4096^3$  simulations.

The seven simulations shown in Figs. 3 and 4 were performed at LANL over the past few months. These currently represent a good fraction of all the high-resolution cosmological simulations in the world (69 billion particles or more). The simulations in Fig. 3 differ only by the spatial scale of the simulation (with corresponding changes in the mass of a particle). The computational volumes span the range of 250 Mpc to 4 Gpc by factors of two, with masses progressing by factors of eight from  $2.44 \times 10^7 M_\odot$  to  $1 \times 10^{11} M_\odot$ . The galaxy halo mass function is shown in Fig. 4, comparing it with the fit derived from our earlier work.

Computer simulations enable discovery. In the words of the Astronomy and Astrophysics Decadal Survey: “Through computer modeling, we understand the deep implications of our very detailed observational data and formulate new theories to stimulate further observations.” The only way to accurately model the evolution of dark matter in the universe is through the use of advanced algorithms on massively parallel computers. On the order of a billion dollars will be invested over the coming years in observational projects probing for signatures of dark matter and dark energy. The return on this investment depends a great deal on having a robust and accurate suite of simulations to interpret these observations in the light of our theoretical models. We have demonstrated our ability to achieve the scale and accuracy necessary using our advanced parallel algorithms and a computer with hundreds of Terabytes of memory and Petascale performance.

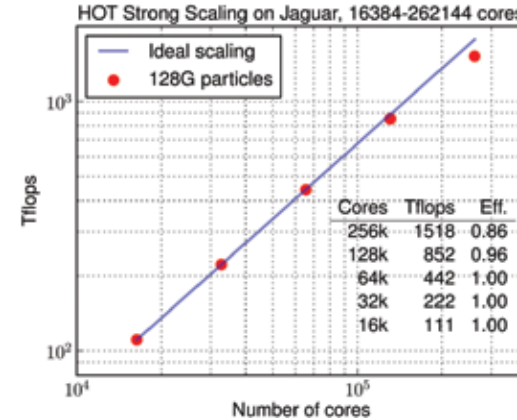


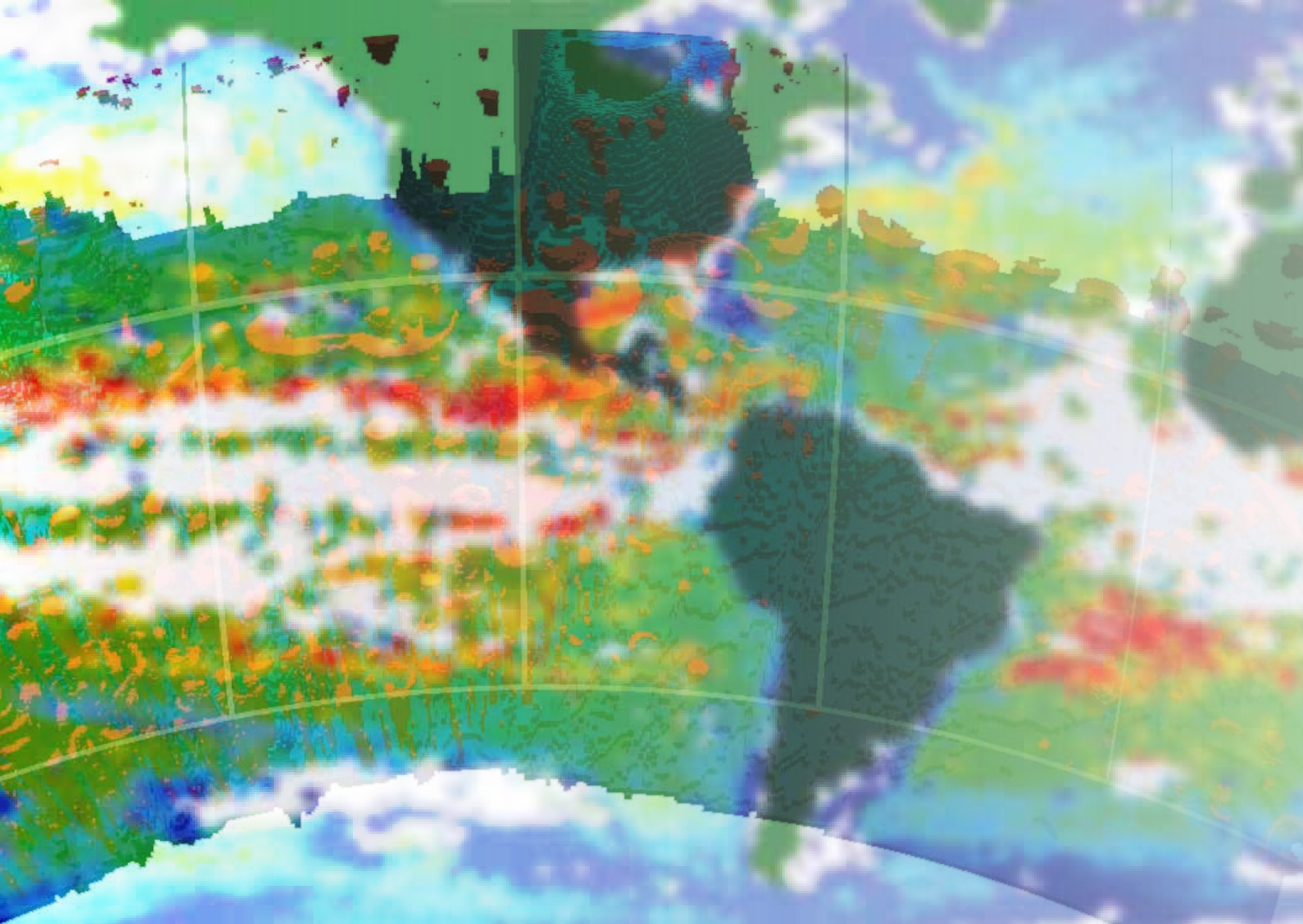
Fig. 4. Scaling on Jaguar at ORNL measured in June 2012.

#### Bibliography

Warren, M.S. and Bergen, B., “The Hashed Oct-Tree  $N$ -body Algorithm at a Petaflop.” Proceedings IEEE Comp. Soc. (2012).

Tinker, J.L., *Astrophys J* **688**, 709 (2008).





# Climate, Atmospheric, and Earth Systems Modeling

LANL is a world-renowned research institution in the area of climate and ocean modeling, stemming from its seminal work on nuclear winter several decades ago which derived from LANL's nuclear weapons activities. LANL scientists are responsible for the ocean and sea ice models in the Community Climate System Model and, as such, are also key contributors to Intergovernmental Panel on Climate Change analyses. The articles in this section highlight new discoveries using state-of-the-art, integrated simulations (glacial response to climate change, understanding unique observations of ocean circulation), new research in the creation of more accurate simulation tools with better modeling capabilities (new parameterizations of sea ice melt ponds, the application of uncertainty quantification), new modeling results for ocean simulations and three-dimensional eddies, the exploration of important new techniques for multi-resolution modeling (a requirement for the understanding of regional climate effects), and seminal work in the understanding of ocean turbulence and eddy formation. These developments are key in the ongoing quest to understand and quantify our changing climate.



# Scale-Aware Sub-Grid Eddy Parametrizations for Ocean Modeling

Qingshan Chen,  
Todd Ringler, T-3

The multi-resolution approach towards global ocean modeling is attractive because it can achieve fine resolutions in regions of interests with manageable computational costs. The success of this approach crucially depends on the access to scale-aware sub-grid eddy parameterizations that can seamlessly transition between different length scales existing on the meshes. Effort is being made to meet this demand, and some initial results are reported here. A scale-aware version of the traditional anticipated potential vorticity method is developed, which can be used on a wide range of multi-resolution meshes without any adjustment of the single non-dimensional parameter needed in the closure. The Gent-McWilliams (GM) closure for mesoscale ocean eddy transport is being studied, and the merits of spatially varying coefficients are evaluated in the context of the Antarctic Circumpolar Current. Results from this study will provide guidelines for the next steps in developing a scale-aware version of the Gent-McWilliams closure.

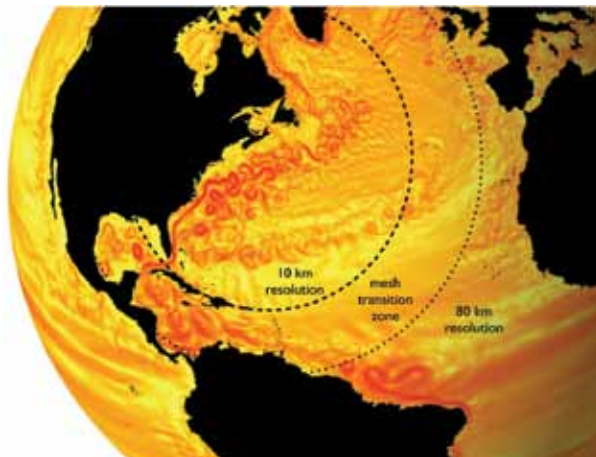
**T**he ocean plays a critical role in the evolution of our climate system because it is involved in the global water circulation and global heat transfer (e.g., equator-to-pole heat transfer). Through its normal or anomalous behaviors, the ocean also directly affects our human society. For example, the Gulf Stream, which is an intensified western boundary current of the subtropic gyre in the Atlantic, transports warm water from the Gulf of Mexico to the northern Atlantic and produces mild winters in Europe. Hence it is no overstatement that, in order to have a decent understanding of how the climate system works, an accurate representation of the ocean is essential. But the latter has been an extremely challenging task, largely due to the wide range of scales present in the ocean and the nearly inviscid nature of the flow.

While unstructured grids have been in use in engineering for many years, the lack of stable numerical schemes suitable for highly rotating geophysical fluid dynamics has precluded adoption of unstructured meshes in the climate community. With unstructured grids, polar singularities that exist on structured grids, such as traditional latitude-longitude grids, are removed, and it is possible to have local refinements on regions of interest and coarse resolutions elsewhere. Model Prediction Across Scales (MPAS) is a joint project between LANL and the National Center for Atmospheric Research (NCAR) to develop finite volume dynamic cores on unstructured grids. Employing the multi-resolution

approach, MPAS-Ocean [1] is now able to fully resolve the mesoscale eddies in regions of interest (see Fig. 1), with computational costs that are manageable by today's supercomputers. With local refinements within a global grid, MPAS-Ocean also provides an appealing alternative for regional climate modeling by allowing regional focus within a global simulation.

Mesoscale eddies (30–250 km) play an important role in the ocean dynamics as they facilitate the transport of heat and other tracers. But to date, long-term climate simulations still cannot afford to resolve the mesoscale eddies and they therefore have to be parametrized. Mesoscale eddy parametrization (MEP) is an essential component of all ocean models. Climate models operating on unstructured multi-resolution grids place a new requirement on the development of MEP schemes, namely they should be able to transition smoothly across all the grid lengths existing on the grid—that is, being scale-aware. We endeavor to develop, evaluate, and implement scale-aware MEP schemes for the next-generation climate models.

To explore the methodology, in the first step we study and generalize the Anticipated Potential Vorticity Method (APVM) [2], which is an eddy closure that is perfectly energy conserving and (potential) enstrophy dissipating. The traditional APVM involves a coefficient that depends on a parameter usually taken as a constant in applications. We employ the 2D isotropic turbulence theory and derive a new formulation for the APVM coefficient that depends on a single scale-invariant parameter. The new formulation is tested in an idealized shallow-water model, first on quasi-uniform unstructured grids [3], and then on variable-resolution



*Fig. 1. A snapshot of the kinetic energy on a 10- to 80-km multi-resolution grid. The eddies along the Gulf Stream are fully resolved, thanks to the high resolution in the northern Atlantic.*



#### Funding Acknowledgments

DOE, Office of Science, Advanced Scientific Computing Research Program, Computational Science Research and Partnership (SciDAC); BER, Regional Climate Modeling Project

unstructured grids [4]. Figure 2 can best demonstrate the scale-aware property of the new formulation. It contains plots of the potential enstrophy (PE—the variance of the potential vorticity) spectral density curves, generated with an empirically determined optimal value for the single parameter of the new formulation, on a set of variable-resolution grids. For each of these variable-resolution grids, the new formulation of the APVM is able to produce a PE spectrum curve that is close to the reference spectrum curve across the entire inertial range. This shows that the new formulation is scale-aware, and robust in the sense that the optimal value for the parameter determined on one grid can be satisfactorily used on other grids as well.

Taking one step forward, we study the GM closure [5,6]. GM is a closure of the mesoscale eddy transport of passive tracers. It is an essential component of almost every ocean model out there, and it is hard to over-estimate the role it has played in ocean modeling. GM is based on the fundamental assumption that eddy transport of tracers should be down the thickness gradient and along, not across, the isopycnal surfaces. The GM closure is far more physically relevant and far more complex than APVM. A scale-aware version of GM, assuming it exists, will necessarily exhibit non-uniformity in space and be attuned to the baroclinic instabilities. We begin our study of GM by seeking answers to the preliminary but important question of whether using spatially varying eddy diffusivities can lead to improved results [7]. We approach this question with a set of controlled simulations using a three-layer isopycnal model of the Antarctic Circumpolar Current (ACC). In this study we take the most direct approach by deriving the eddy diffusivities from a high-resolution reference simulation and then feeding the derived diffusivities to low-resolution simulations. The results demonstrate that using spatially varying eddy diffusivities remedies the major deficiency of the extended GM closure [8]. It also leads to improved results with regard to the thickness and potential vorticity

(PV) profiles and the volume fluxes. Results from this study will provide guidelines in developing a spatially varying GM coefficient and ultimately a scale-aware version of the GM closure.

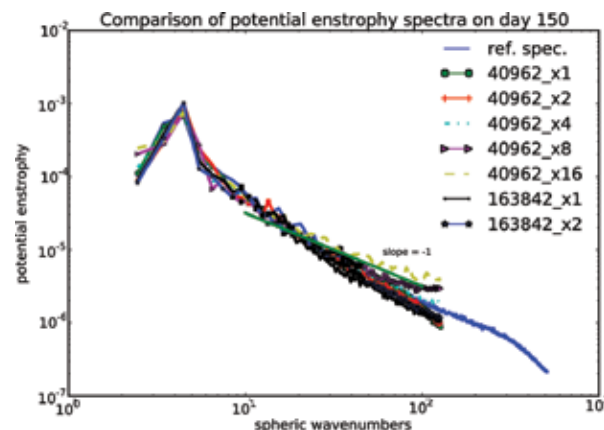


Fig. 2. The potential enstrophy spectrum curves generated by a single value (0.0020) for the APVM parameter on a set of multi-resolution grids. In the grid designation  $n1\_xn2$ ,  $n1$  denotes the number of grid cells, and  $n2$  denotes the ratio between the coarsest and finest resolutions, which can be up to 16 for our study.

- [1] M. Petersen et al. *ADTSC Science Highlight* in press 50, (2013)
- [2] Sadourny, R. and C. Basdevant, *J Atmos Sci* **42**(13), 1353 (1985).
- [3] Chen, Q. et al., *Mon Weather Rev* **139**(8), 2614 (2011).
- [4] Chen, Q. et al., *Mon Weather Rev* **140**(9), 3127 (2012).
- [5] Gent, P.R. and J.C. McWilliams, *J Phys Oceanogr* **20**(1), 150 (1990).
- [6] Gent, P.R. et al., *J Phys Oceanogr* **25**(4), 463 (1995).
- [7] Chen, Q. et al., *Ocean Model* submitted (2012).
- [8] Ringler, T. and P. Gent, *Ocean Model* **39**, 125 (2011).

# Uncertainty Quantification and Ocean Model Analysis

James R. Gattiker,  
Sham Bhat, CCS-6;  
Matthew W. Hecht, CCS-2

Uncertainty quantification (UQ) methods are a tool for analysis of the performance of simulation models. This research summary describes how the application of UQ tools to the POP2 ocean model can be used to quantify model improvements, to understand where next-generation models can be improved, and to determine the settings of parameters and uncertainty in model applications.

Uncertainty quantification (UQ) is concerned with learning about model performance from the comparison between model output and accepted performance targets. This research uses LANL's Bayesian analysis tool GPM/SA (Gaussian Process Models for Simulation Analysis) [1] to study the Community Earth System Model's ocean component, the LANL-developed POP2 [2]. Two products of the analysis are discussed: (1) a method for quantitatively assessing model performance for qualification of model improvements, and (2) an analysis of distributions on the parameters of key model settings in the face of conflicting measures of model performance.

Complex physical models often use parameterizations to overcome limitations. The Gent-McWilliams isopycnal transport and mixing scheme (GM) [3] corrects for the transport by ocean eddies that cannot

stratification-keyed weakening coefficient called tapering. The specific application goal in this project is to apply UQ analysis to the uncertain parameters of GM, determining its quality and best settings.

Ocean model analysis challenges include: (1) sparse Earth system observations, especially for the deep ocean, making direct model-observation comparison problematic; and (2) the ocean state contains long time scales, so the amount of computation required to reach a statistical steady state is daunting. To address these challenges, a test-bed ocean configuration is adopted whose scope and computation are considerably reduced from the full Earth system. The "channel model" is designed to study eddy-induced transport, mimicking the Antarctic Circumpolar Current of the Southern Ocean. The reduced scope makes the computation of high-resolution simulations explicitly resolving eddies and their mixing effects feasible. A high-resolution simulation can then be taken as "truth" for the purposes of studying the GM parameters. Figure 1 shows the channel model's sea surface temperature in the high-resolution result, as it is to be compared to an ensemble of low-resolution model runs. The ensemble is designed to span the plausible ranges of the GM parameters. This ensemble allows the construction of a fast statistical emulator of the model's response that can be used to support inferences in the comparison, including parameter distributions and measures for how well the model fits the dataset.

A key element in comparing the simulations to their targets is the specific set of metrics to be used for the comparison. The metrics define the question we are asking of the models and the analysis. Figure 2 shows four metrics used for analysis, from a larger set of measures suggested by domain experts. Each member of this set is an extracted property versus depth, averaged horizontally over the simulation domain, and over a time window of five years of simulation, after equilibration. Temperature and salinity profiles show direct impacts of transport and mixing, the density in turn is dependent on these, and the vertical heat transport summarizes the movement of heat. Because we are

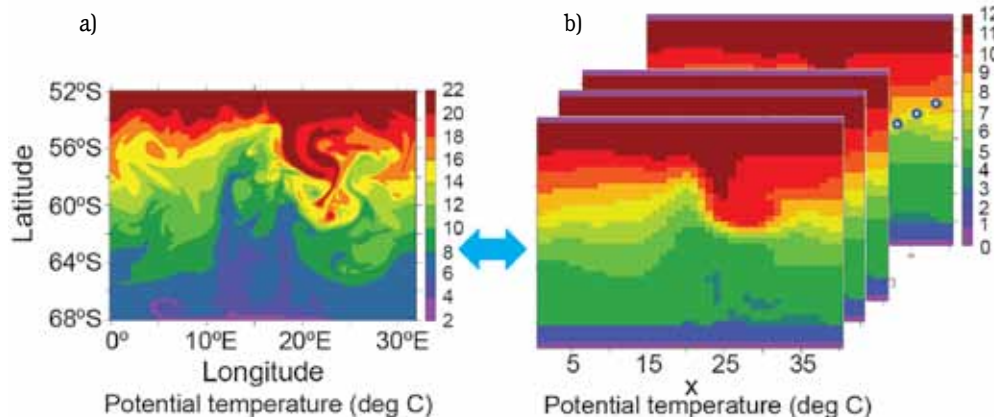


Fig. 1. Uncertainty quantification compares a) a reference high-resolution solution to b) many examples of low-resolution simulations at various settings of uncertain parameters.

be resolved in the current generation of climate simulations, run at grid resolutions comparable to our 0.8-degree grid simulations (Fig. 1b). The GM parameterization is arguably the most significant parameterization in ocean models used for long climate simulations, greatly improving the representation of ocean heat transports [4]. The current GM formulation relies on two uncertain parameter settings, an overall magnitude and a

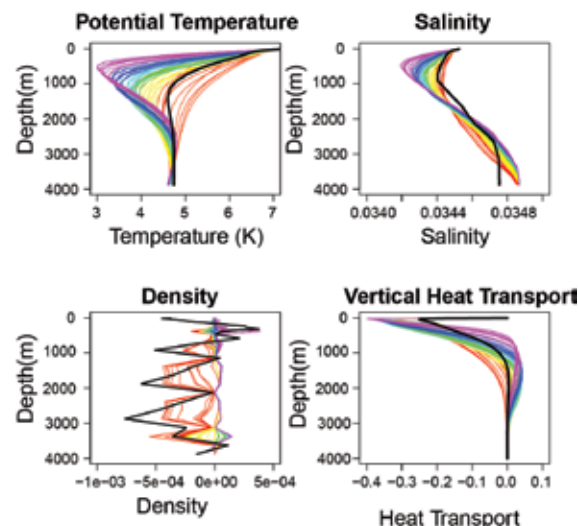


Fig. 2. Four metrics for comparison: Potential Temperature, Salinity, Density (residual from simulations' mean), and Vertical Heat Transport. Black is high-resolution, color is a composite related to the sum of parameters for presentation.

attempting to approximate with the GM scheme the effects of computationally complex fluid dynamics, it is not expected that we can meet the resolved simulation results in every detail. Qualitatively, these metrics appear in some cases to be very informative, and in other cases show difficulty.

UQ analysis accounts for uncertainty in the emulated model response, uncertainty in the parameters controlling the match between the low-resolution model and its target, uncertainty in the general quality

of fit between these models, and uncertainty regarding irreducible structural differences. These aspects cannot be considered in isolation, they are all part of the same problem of comparing simulations to a reference target, in this case the low-resolution parameterized model to the high-resolution target. The various interdependent aspects of uncertainty must be considered even when the focus is on one particular aspect. The UQ approach uses Gaussian process emulator models, an additive Gaussian process for structural discrepancy, and parameters for overall model fit expressed as a target precision. The result of analysis is a joint probability distribution of parameters, both those in the statistical model and those controlling the behavior of GM.

In our first goal to demonstrate the capability to qualify model improvements, we are interested in the summary measure of the ability of the model with GM to fit the target. We compare this measure for an older one-parameter GM implementation, the two parameters of the current GM, and the value of considering a third parameter related to the shape of the GM tapering. The improved model (one parameter to two parameters) is clearly quantifiable, while

exposing the addition of a third parameter does not represent an advance in model fit. In this case the additional parameter isn't independently contributing to quality fit with respect to the metrics.

The complete target of a UQ analysis is to understand what parameter settings are implied by the data, or perform "inverse" analysis—again, taking into account the various sources of uncertainty. Each metric results in a distribution on parameters, shown in the colored curves in the axes of Fig. 3. A novel aspect of this analysis is the methodology used to combine these various results. The composite distribution should attempt to unify the results where they are feasibly similar, but when there is unresolvable conflict should admit appropriate uncertainty. An appropriate model for this is a Bayesian hierarchical model across relevant metrics with a prior preferring consistency. Figure 3 shows how unification proceeds while adding indicators. The top axis shows that the implications of the temperature and salinity metrics are consistent. The middle axis shows that although the results of density are not independently identical, that information can be unified with the previous result. However, vertical heat transport represents information in conflict, hence its inclusion results in a large increase in uncertainty. In the problem context, we know that GM cannot improve all possible measures. The result of the UQ analysis communicates to the domain expert where these various metrics can be improved by GM consistently and where there is substantial discrepancy, providing valuable input to the next cycle of model improvements. Today, this information assists in understanding how to set the values of the current generation of applied models to achieve the best performance.

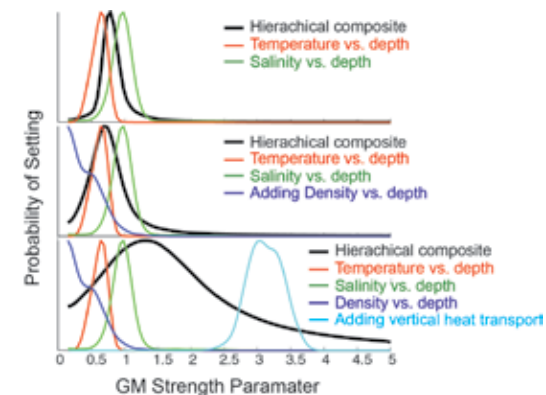


Fig. 3. Summary of parameter distributions with respect to different metric sets.

[1] Higdon, D. et al., *J Am Stat Assoc* **103**(482), 570-583 (2008).

[2] Smith, R. et al., "The Parallel Ocean Program (POP) Reference Manual," LANL Technical Report LA-UR-10-01853 (2010).

[3] Gent, P.R. and J.C. McWilliams, *J Phys Oceanogr* **20**, 150 (1990).

[4] Gent, P.R. et al., *J Clim* **11**, 1287 (1998).



# Level-Ice Melt Ponds in the LANL Sea Ice Model, CICE

Elizabeth C. Hunke, T-3;  
David Hebert, Naval Research  
Laboratory, Stennis Space  
Center;  
Olivier Lecomte, Université  
Catholique de Louvain,  
Belgium

In a new melt pond parameterization developed for the LANL sea ice model CICE, the ponds evolve according to physically based process descriptions, accounting for the topography of the sea ice. Realistic model hindcast simulations (1958–2007) are used to explore the interactions of physical mechanisms that affect the evolution of ponds and sea ice albedo, highlighting the importance of snow melt processes and precipitation rates for trends in ice volume. Various feedback mechanisms also come into play, including a new level-ice-pond feedback that enhances the rate of sea ice thinning.

Surface characteristics determine the surface energy balance of sea ice. Sea ice volume is highly sensitive to the thermodynamic fluxes that determine this balance, of which short- and long-wave radiation are critical components in summer. Summer melt ponds, pools of melted snow and ice that collect in depressions on the ice surface, are relatively dark in color and can lower the surface albedo considerably from the relatively high values associated with snow cover and bare ice. The surface albedo continues to decrease as more melt water collects on the ice, increasing solar absorption and further melting the ice and snow, an important albedo feedback process. The formation, evolution, and disappearance of melt ponds are governed by complex processes, including interactions with the existing snow layer, drainage rates through permeable sea ice, episodic refreezing, and considerations of ice topography, making detailed melt pond modeling a daunting task.

A new melt pond parameterization has been developed for the LANL sea ice model, CICE [1]. The ponds evolve according to physically based process descriptions, assuming a depth-area ratio [2] for changes in pond volume. A novel aspect of the new scheme is that the ponds are carried as tracers on the level (undeformed) ice area of each thickness category, thus limiting their spatial extent based on the simulated sea ice topography. This limiting is meant to approximate the horizontal drainage of melt water into depressions in ice floes. Infiltration of the snow by melt water postpones the appearance of ponds and the subsequent acceleration of melting through albedo feedback, while snow on top of refrozen pond ice also reduces the ponds' effect on the radiation budget. Other simulated melt pond processes include collection of liquid melt water and rain into ponds, drainage through permeable sea ice or over the edges of floes, and refreezing of ponds [3].

Melt ponds first appear at southerly latitudes in spring, moving north as the melt season progresses (Fig. 1). Ponds form quickly and are

widespread initially, then pool into low topographic features and begin to drain through permeable ice within a few weeks. Pond area then slowly increases due to continuing snow and ice melt until the ponds' upper surfaces begin to refreeze. Because of variations in topography and permeability, we find deeper ponds on thicker, more deformed ice, and these are the last to freeze over in autumn.

Interannual variability is significant throughout the simulations, driven by three non-climatological forcing fields: air temperature, humidity, and wind velocity. Feedback processes are important for strengthening the variability. These include the ice-albedo feedback, in which changing ice surface characteristics enhance or reduce melting, and the ice-ocean albedo feedback in which reduced ice cover reveals the darker ocean surface, allowing greater heat absorption and additional ice melt; the ice-ocean albedo feedback also works in reverse, with increased ice area leading to ocean cooling and further ice formation. In this study we discovered a level-ice-pond feedback mechanism that has not been described previously, in which thinning ice has more level surface area available to be covered in ponds, enhancing thinning.

Sensitivity tests reveal that the snow simulation is critical, because the volume of snow deposition and rate of snow melt largely determine the timing and extent of the simulated melt ponds. Nevertheless, compensating effects moderate the model's sensitivity to precipitation changes. For instance, infiltration of the snow by melt water postpones the appearance of ponds and the subsequent acceleration of melting through albedo feedback, while snow on top of refrozen pond ice also reduces the ponds' effect on the radiation budget. With less snow insulating the ice (in more recent years), we would expect greater conduction and sea ice growth in winter, but increased radiation penetrating the ice slows growth and/or enhances summer melting, leading to thinner sea ice overall.

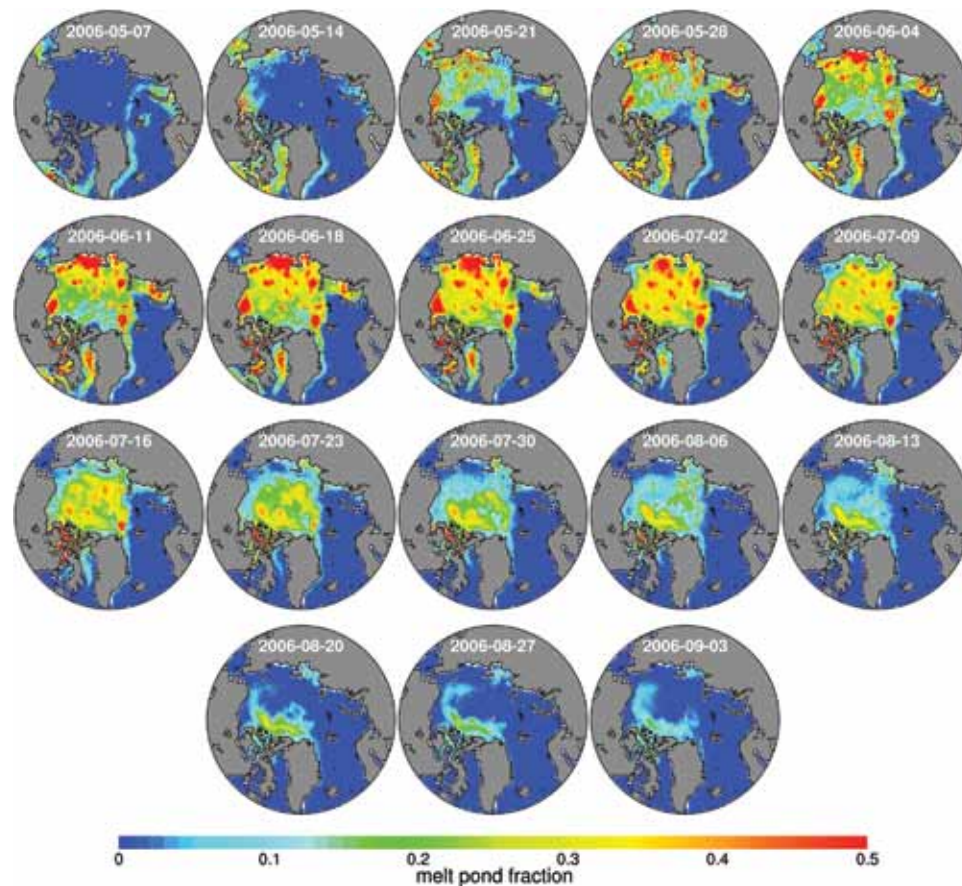
#### Funding Acknowledgments

DOE, Office of Science, BER, Earth System Modeling and Regional and Global Climate Modeling programs; Naval Research Laboratory, DISTANCE program; European Commission's 7th Framework Programme, COMBINE project

#### Special Thanks

This modeling venture originated as a student project at the IARC Summer School on "Modeling of the Arctic Climate System," International Arctic Research Center, Fairbanks, Alaska, May 2011.

Fig. 1. Ponded fraction of ice area, simulated for the Arctic in 2006.



[1] Hunke, E. et al., "Level-ice Melt Ponds in the Los Alamos Sea Ice Model, CICE," LA-UR 12-21874; *Ocean Model*, in revision (2012).

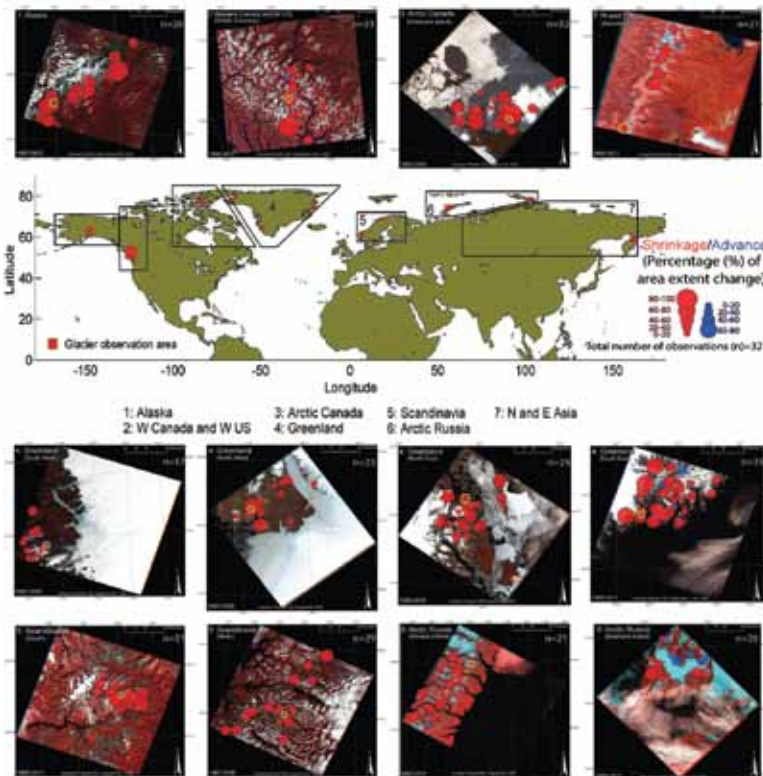
[2] Holland, M.M., et al., "Improved Sea Ice Shortwave Radiation Physics in CCSM4: The Impact of Melt Ponds and Aerosols on Arctic Sea Ice," LA-UR 11-10223; *J Clim* **25**, 1413 (2012).

[3] Flocco, D. et al., "Impact of Melt Ponds on Arctic Sea Ice Simulations from 1990 to 2007," LA-UR12-21872; *J Geophys Res* doi:10.1029/2012JC008195, in press (2012).

# Glaciers' Response in a Warming Climate

Sebastian H. Mernild, CCS-2

Fig. 1. Satellite-derived area changes of 321 glaciers and ice caps in the Pan-Arctic. Changes are shown as rates during the observation period from approximately mid-1980s to present (the period varies between regions, and for N and E Asia the observation period was 1999 to present). The data were divided into seven first-order glaciated regions: (1) Alaska, (2) W Canada and W US, (3) Arctic Canada, (4) Greenland, (5) Scandinavia, (6) Arctic Russia, and (7) N and E Asia. Three of the first-order regions were divided into sub-regions illustrated on the Landsat images, indicating in total 12 sub-regions. Red circles show (GIC) shrinkage and blue circles advance (%). Circles with green margin show examples of GIC margin and area changes illustrated in Fig. 2. Background satellite images are from Landsat 5 TM.



The latest decade (2001–2010) was the warmest of the instrumental record, and more so at high latitudes—it is likely that this decade has been warmer than any other decade within the last thousand years. In the Arctic, for example, most observed glaciers and ice caps (GIC) shrank in area, indicating that glaciers and ice caps have lost an arithmetic average of one-fifth of their area since the mid-1980s. Overall, this shrinking area follows the observed mean global glacier and ice cap mass trend towards negative balances; however, the mass balance from the most recent pentad (2006–2010) shows more moderate, although still large, losses.

Glaciers and ice caps (this includes all glaciers except the Greenland Ice Sheet and the Antarctic Ice Sheet) are tracers of climate changes because air temperature and snowfall control their surface mass balance [1,2]. At high northern latitudes, the temperature rise of recent years has been more pronounced than the global average and around twice the global average for the past 100 years [3]. For example for Greenland, based on long-term temperature observations, the last decade was not only the warmest since 1890 but it also had the highest number of extreme warm years, higher by around 50% than the number in the warm 1930s and 1940s [4].

The circumpolar Arctic region contains half of the estimated global glacier and ice cap surface area and two-thirds of its volume [5]. Even though GIC account for less than 1% of all the water on Earth that is bound

in glacier ice, their increasing retreat and mass loss may dominate the glacial component of the global sea-level rise of the past century [6,7]. Analyses show that glacier and ice cap mass losses are currently raising the mean global sealevel by approximately 1 mm sea level equivalent (SLE) per year [7,8], which is broadly similar to the combined contributions from the Greenland Ice Sheet and the Antarctic Ice Sheet [8,9].

Throughout the approximately 25 years of satellite coverage, the Pan-Arctic GIC's have faced widespread non-uniform shrinkage, where only 8%, 26 out of the 321 observed glaciers and ice caps, advanced in area [10] (Fig. 1). As an example, these non-uniform area changes are illustrated for 12 individual glaciers and ice caps (for both minor and major, Fig. 2), where 10 out of 12 showed retreat. On a sub-regional scale, half of the 12 regions showed retreat for all observed glaciers and ice caps, whereas 5% of the glaciers and ice caps advanced in Novaya Zemlya, 9% in SE Greenland, 10% in Kamchatka, 13% in W. Canada and the W. US, 37% in SW Greenland, and 42% on Bolshevik Island. The glacier and ice cap area changes on Bolshevik Island are not described in detail in the literature, therefore it is not yet possible to know whether the advancement is due more to positive net mass balance (climatic response) or to surging activities (climate-dynamic response). It is notable, however, that all the glaciers and ice caps on Bolshevik Island are facing north. Advancing glaciers and ice caps are, for all sub-regions, predominantly facing north (85%), and subject to dynamic response to changes in positive mass balance and climate. Surging activity is present and influences a real extent in Arctic Russia and Kamchatka. For example, for Novaya Zemlya 5% of the observed glaciers and ice caps advanced, where an identical percentage of glaciers and ice caps are known to surge [11], with 32 potential surge-types identified out of 692 glaciers and ice caps on the Novaya Zemlya archipelago.

For the compiled glacier and ice cap data set, the arithmetic mean relative area change was  $21 \pm 1\%$  from the mid-1980s to the present (here and following, the error term is stated as plus or minus one standard error). On the regional scale, Alaska faced an average shrinkage of



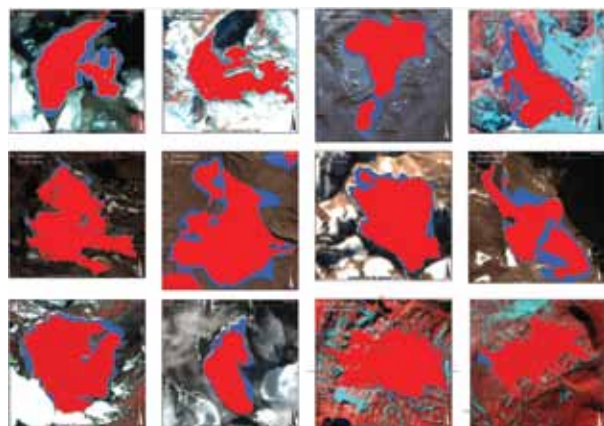
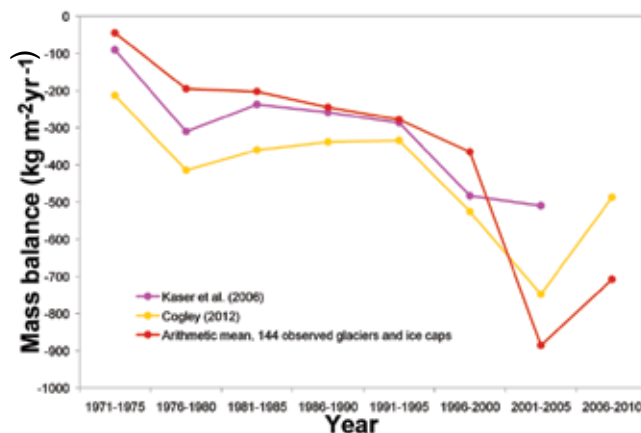


Fig. 2. Examples of satellite-derived margin location and area changes of 12 chosen glaciers and ice caps (for both minor and major glaciers and ice caps, one for each sub-region) for the period mid-1980s (blue) to present (red) estimated from Landsat images. The location of the GIC is shown in Fig. 1.



40±4%, Arctic Canada 35±4%, N and E Asia 23±3%, Scandinavia 21±2%, Greenland 20±2%, W. Canada and the W. US 12±3%, and Arctic Russia 12±2% (Fig. 1). For the Canadian Rocky Mountains, an area shrinkage of 15–25% (approximately 1950–2000) was computed [12,13], but a direct comparison to previous studies cannot be made due to the uneven observation periods. Overall, this

shrinking trend follows the observed mean global glacier and ice cap mass balance trend towards negative balances (based on data from 144 observed glaciers and ice caps, Fig. 3). The mass balance from the most recent pentad (2006–2010) shows more moderate, although still large, losses.

Historically, the representation of shrinking and advancing glaciers and ice caps conditions has been either non-existent or limited to Pan-Arctic regions; however, satellite and aerial observations from SE Greenland go back to the 1930s [14]. Simultaneously mapping the shrinking and advancing behavior of Pan-Arctic glaciers and ice caps provides quantitative insight into the climate impacts on the cryosphere. For the last approximately 25 years shrinking of land-terminating glaciers and ice caps has been

Fig. 3. Estimated global average glacier and ice cap mass balance at pentadal intervals (1971–2010) from published estimates and from the arithmetic mean of 144 glaciers and ice caps (red color).

documented in high-latitude regions, covering a variation in loss rates from 40% in Alaska to 12% in Arctic Russia. To understand the glacier and ice cap response to climate change a new model study has just been initiated (Fig. 4) from which we will learn about the glacier and ice cap mass balance behavior across the Northern Hemisphere (north of 25° N) for the last four decades, clarifying the glacier and ice cap response in a warming climate and the temporal and spatial contribution of glacial mass loss to global sea-level rise.

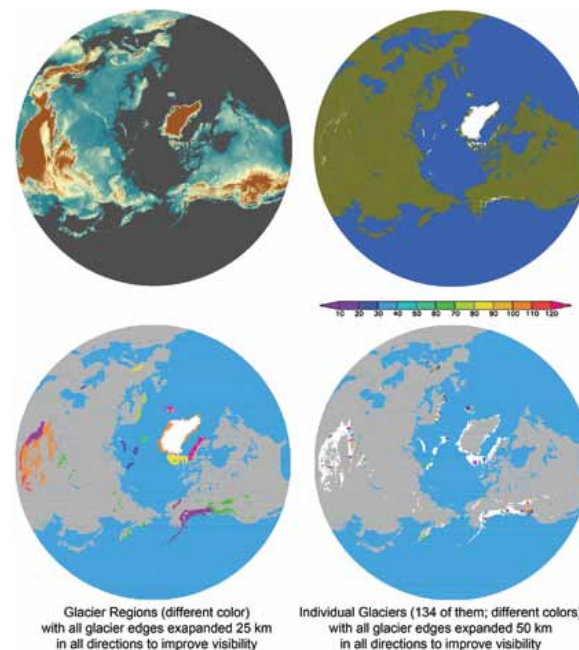


Fig. 4. Northern Hemisphere simulation domain: (upper left) topography (shaded from blue color [low elevations] to brown color [high elevations]); (upper right) surface characteristics (white color is glacier ice, green is grassland, forest, urban, etc., and blue is ocean and lakes); (lower left) different glacier and ice cap regions (following the IPCC AR5 regional division); and (lower right) locations of glaciers and ice caps where mass balance has been observed (within the simulation domain 134 glaciers and ice caps have been observed periodically; all individual glacier and ice caps are marked with a colored dot).

[1] Meier, M.F., *The Quaternary of the United States*, Eds H.E. Wright, D.G. Frey, Princeton University Press, New Jersey, 795 (1965).

[2] Kaser, G. et al., *Geophys Res Lett* **33**, 1 (2006).

[3] Hansen, J. et al., *Rev Geophys* **48**, RG4004 (2010).

[4] Mernild, S.H. et al., "Coastal Greenland Air Temperature Trends", *Int J Climatol*, in review (2012).

[5] Radić, V. and R.J. Hock, *Geophys Res* **115**, F01010 (2010).

[6] Leclercq, P.W. et al., *Surv Geophys* **32**, 519; doi:10.1007/s10712-011-9121-7 (2011).

[7] Radić, V. and R. Hock, *Nature Geosci* **4**, 91 (2011).

[8] Cogley, J.G., *Ann Glaciol* **50**, 96 (2009).

[9] Meier, M.F. et al., *Science* **317**, 1064 (2007).

[10] Mernild, S.H. and J. K. Malmros, *The Cryosphere Discussion* **6**, 4417 (2012).

[11] Grant, K.L. et al., *J Glaciol* **55**(194), 960 (2009).

[12] Luckman, B.H. and T. Kavavagh, *Ambio* **29**, 371 (2000).

[13] DeBeer, C.M. and M.J. Sharp, *Ann Glaciol* **46**, 215 (2007).

[14] Bjork, A.A. et al., *Nature Geoscience* **5**; doi:10.1038/NCEO1481 (2012).

# The Arctic Terrestrial Simulator: Modeling Permafrost Degradation in a Warming Arctic

Ethan Coon,  
Rao Garimella,  
Gianmarco Manzini,  
J. David Moulton,  
Daniil Svyatskiy, T-5;  
Markus Berndt, CCS-2;  
Scott Painter, EES-16

The terrestrial Arctic has been a net sink of carbon for thousands of years, but warming trends suggest this may change. Predictive simulation of the fate of this carbon is critical for understanding feedback effects between the terrestrial Arctic and climate change. Toward this end, a flexible, versatile modeling capability called the Arctic Terrestrial Simulator (ATS) is being developed. Here we introduce the modeling challenges of the problem and discuss computational advances required for predictive simulation, including the development of a multiphysics simulation framework and meshing infrastructure.

Researchers at LANL are developing a new high-performance simulation tool—the ATS—to study the effects of climate change on the release of massive amounts of carbon currently frozen in the Arctic.

It is estimated that approximately 1,700 billion metric tons of organic carbon are frozen in the Arctic permafrost [1]. The tipping point for the release of this carbon in the form of greenhouse gases due to permanent degradation of the permafrost is unknown [2]. Degradation of the permafrost, consequent evolution of the topography, the resulting restructuring of the drainage networks, and their effect on the vegetation are poorly understood. Computer modeling is a key tool in untangling these complex couplings to understand the evolution and potential feedback of the Arctic and subarctic landscapes on the global climate system.

LANL researchers are developing ATS as a ground-breaking modeling capability to simulate freeze/thaw cycles in the arctic soil, snow melt and runoff, subsurface flow due to infiltration and ice melt, soil subsidence due to the melting ice, and biogeochemical

processes in the vegetation layer [3]. This is a computationally challenging problem because of the number and complexity of the controlling physical processes, and the strong coupling between these processes. Necessarily fine spatial and temporal resolution, in conjunction with nonlinearities due to the freeze/thaw cycle, pose additional difficulties for the simulator. Finally, handling the evolving topography places demanding constraints on the mesh infrastructure and discretization strategies.

The ATS is based on Amanzi [4], the open-source flow and reactive-transport simulator being developed by the Advanced Simulation Capability for Environmental Management (ASCEM) program. ASCEM

is a multi-lab program developing a state-of-the-art approach and open-source integrated toolset for the assessment and management of DOE legacy waste sites. The combined arctic processes in ATS and multiphysics capabilities in Amanzi are called Amanzi-ATS.

Amanzi-ATS is built with flexibility and modularity as its main design principles. Individual problems simulated using Amanzi-ATS may require different processes, or may require coupling different processes in different ways, to correctly capture the essential physics. To accomplish the flexibility required, processes are combined in a tree structure that characterizes how Process Kernels (PK) are coupled (see Fig. 1 for a schematic of a possible PK layout for the full Arctic system). PKs are coupled via multi-process coordinators (MPC), which can encapsulate weak (operator-split) or strong (fully implicit) coupling. At any node in the PK tree, a time-integration scheme and nonlinear solver for the coupled subsystem consisting of all PKs below that node in the tree can be automatically formed and solved. This automated coupling allows complex systems to be built and tested quickly and easily. If needed, special couplers can be written that automate most of the process but allow the inclusion of extra coupling terms for preconditioners or solvers.

This PK tree makes it trivial to include, or exclude, individual processes in a given simulation. This is important for two reasons. First, it makes testing individual processes easier. Tests are built up hierarchically, testing first one process, then two coupled processes, then multiple processes together. This builds confidence in the code. Second, it makes evaluating scientific hypotheses easier. Swapping PKs enables different physics to be included, or excluded, determining whether a given process is important to the question of interest.

Amanzi-ATS also has a sophisticated parallel mesh infrastructure to manage the multiple meshes required and then deform those meshes. These meshes are typically formed from data, including topography,

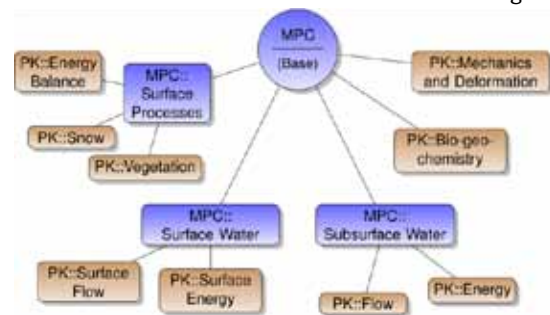


Fig. 1. A schematic of the PKs and coupling needed for permafrost simulation. Brown nodes indicate physical process kernels, which describe one process. Blue nodes MPC, which couple the process kernels below them. To enable the tree structure, MPCs are also themselves PKs, presenting an interface for describing the coupled system as one PK to the MPC above it. The top level (Base) MPC manages the coupling of each subsystem and advances the entire simulation.

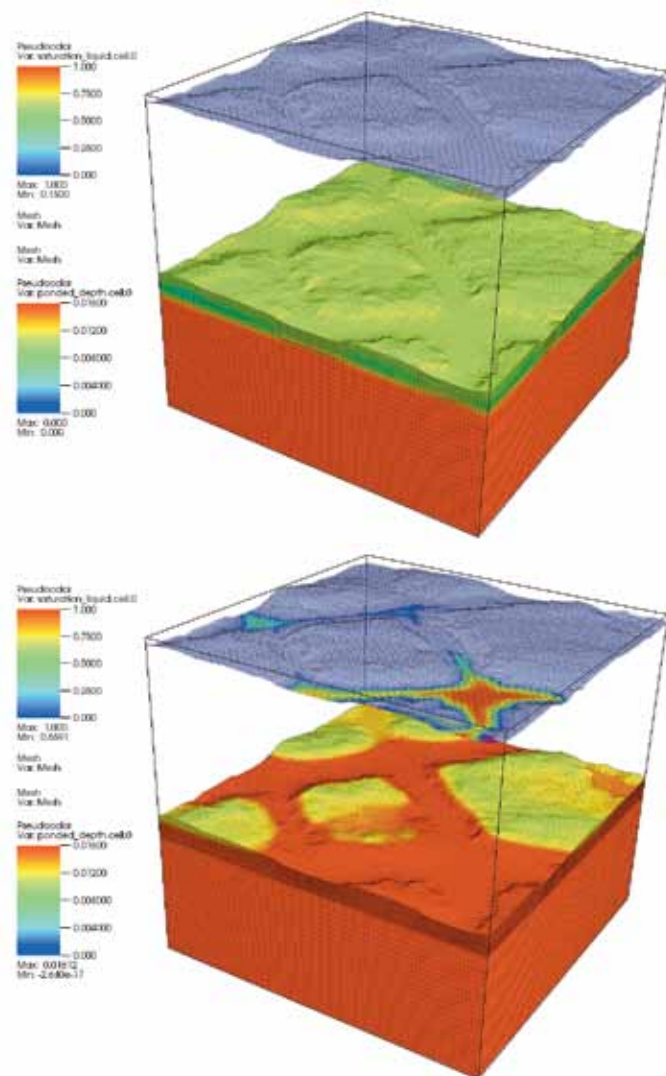


Fig. 2. A demonstration of coupled surface/subsurface flow for high-centered polygons at Barrow, Alaska. In these visualizations, the surface mesh has been raised above the subsurface mesh. Rain falls on the surface of the domain, with initial water table shown (top). The rainfall infiltrates the subsurface until the water table rises above the surface in some regions, where it pools (bottom). The meshes are formed and managed by advanced mesh infrastructure, while coupling is automated by the multiphysics capabilities of Amanzi-ATS.

from sites such as Barrow, Alaska. The mesh infrastructure in Amanzi-ATS wraps the LANL-developed MSTK mesh framework and provides additional functionality tailored specifically for the processes in the Arctic simulations. The process kernels in Amanzi-ATS can query the mesh for entity geometry, topology, and connectivity. In addition, entity sets based on geometrically defined regions are available—these are used for assigning initial conditions, boundary conditions, or material properties. The mesh infrastructure in Amanzi-ATS is capable of reading very large 3D meshes in parallel and weaving them together for the subsurface flow and freeze/thaw simulations. The top surface of the meshes on each partition is extracted and then woven together, forming a parallel surface mesh for overland flow simulation. Finally, Amanzi-ATS also has the ability to deform meshes as driven by a soil subsidence model resulting from ice melt. The subsidence model computes a set of nodal displacements, and the mesh infrastructure attempts to deform the mesh while maintaining the validity of elements.

Amanzi-ATS simulations have been conducted on up to 10,000 processors on LANL institutional clusters. These simulations include the coupled overland and subsurface flow shown in Fig. 2 and cyclic freeze/thaw problems on geological domains in the Arctic.

During the next year, Amanzi-ATS will be developed further to include subsidence, snow, and biogeochemical processes, and to couple these processes. These advances will test both the multi-physics framework and the mesh infrastructure highlighted here. As this predictive simulation capability matures, it will be used to study realistic what-if scenarios of permafrost degradation and greenhouse gas release.

- [1] Tarnocai, C. et al., *Global Biogeochem Cy* **23**, GB2023 (2009).
- [2] “Climate Change 2007: Synthesis Report,” *Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, **104** IPCC (2007).
- [3] Painter, S. et al., “Modeling Challenges for Predicting Hydrologic Response to Degrading Permafrost,” *Hydrogeology J*, in press (2012).
- [4] Moulton, J.D. et al., “Amanzi et al., the Multi-Process High Performance Computing Simulator,” LANL Technical Report LA-UR 12-22193; ASCEM-HPC-2011-03-1 (2012).



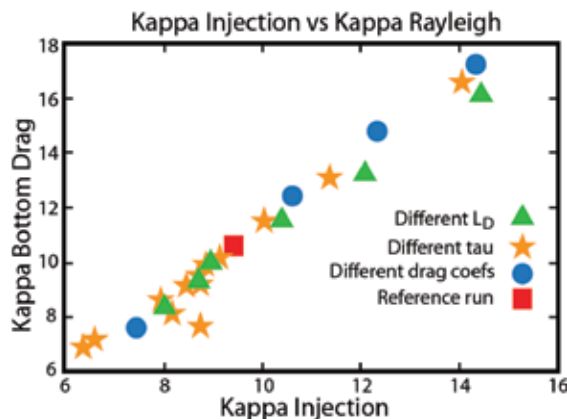
# On a Turbulent Energy Transfer Scale in Ocean Turbulence

Balu T. Nadiga, CCS-2;  
David N. Straub, McGill  
University

The budget of mechanical energy that goes into ocean circulation at large scales is not well understood. This is due to the contrasting nature of turbulence at large scales—rotating and stably stratified balanced turbulence wherein energy cascades to larger scales—and that at smaller scales—unbalanced turbulence wherein energy cascades to smaller scales. In this research program, we systematically investigate the interaction of turbulence across these asymptotic large-scale and small-scale regimes to better constrain the energy budget. The finding we highlight here is related to the extant phenomenological picture of large-scale turbulence, one aspect of which is that the turbulent cascade of kinetic energy into the gravest vertical mode occurs at a characteristic fixed scale, after which that energy cascades up to larger horizontal scales. In contrast, we find that in oceanographically relevant settings, this turbulent energy transfer scale is variable and scales well with the scale of large-scale energy dissipation.

As much as 90% of the energy that goes into the ocean is dissipated in the top 100 meters. The remaining 10%—that is mainly input on the large scales—drives the interior ocean circulation, a crucial component of the climate system [1]. Consequently, an understanding of the energy budget of the latter component is important from the climate perspective.

In one study of this research program, to better understand the energy budget, we conducted numerical simulations of a prototypical wind-driven ocean circulation problem that spans the oceanographically relevant range of values for a number of parameters (Rossby deformation radius that measures stratification, amplitude of wind-stress, and the coefficient of large-scale dissipation) [2]. A robust feature of all of these simulations is the correspondence in scale between that at which turbulent energy is injected into the gravest vertical mode and that from which energy is removed from this vertical mode by bottom drag. A surprisingly good linear relation, with a slope near unity, is evident (Fig. 1).



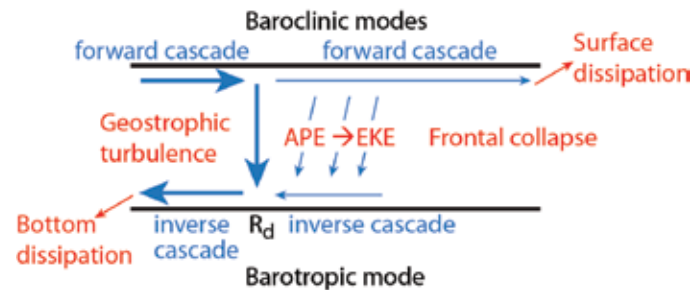


Fig 2. Presently accepted phenomenology of rotating stably stratified turbulence. Barotropic mode refers to the gravest vertical mode; baroclinic modes refer to other vertical modes. According to this picture, turbulent energy is injected into the gravest vertical mode at a fixed scale related to the Rossby deformation radius ( $R_d$ ), which then cascades to larger horizontal scales before being dissipated by bottom friction [1].

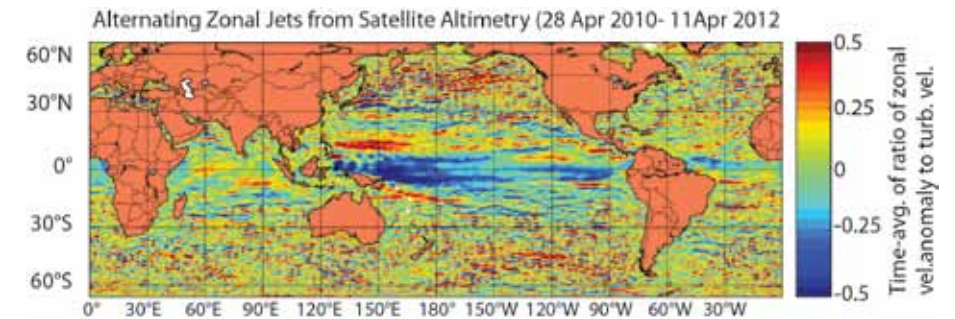


Fig. 3. A two-year average of the ratio of zonal velocity anomaly to turbulent velocity using satellite altimetry (integrated multi-mission ocean altimeter data for climate research and AVISO datasets).

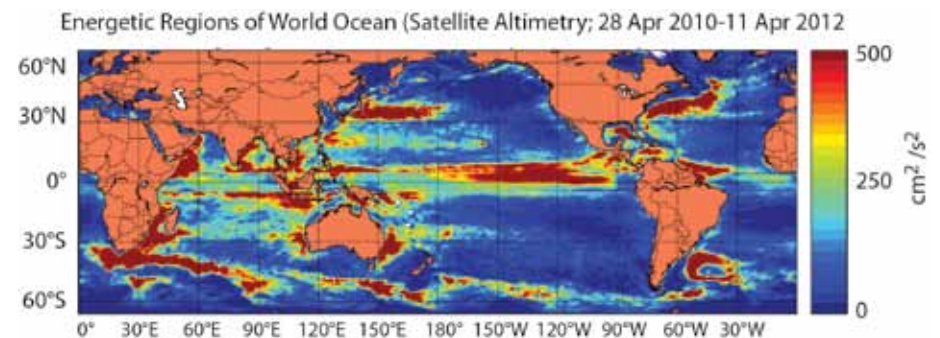


Fig. 4. A spatial map of turbulent kinetic energy for the same period.

- [1] Ferrari, R. and C. Wunsch, *Annu Rev Fluid Mech* **41**, 253 (2009).
- [2] Straub, D.N. and B.T. Nadiga, "Zonal Jets and Energy Cascades in the Baroclinic Ocean Double Gyre Problem," *AGU Fall Meeting Abstracts*, **A1506** (2011).
- [3] Nadiga, B.T., *Geophys Res Letts* **33**, L10601 (2006).
- [4] Nadiga, B.T. and D.N. Straub, *Ocean Model* **3**, 257 (2010).

# Statistical Mechanics of the Geometric Control of Flow Topology in 2D Turbulence

Balu T. Nadiga, CCS-2;  
Peter N. Loxley, T-5

We apply the maximum entropy principle of statistical mechanics to 2D turbulence in a new fashion to predict the effect of geometry on flow topology. We consider two prototypical regimes of turbulence that lead to frequently observed self-organized coherent structures. Our theory predicts bistable behavior that exhibits hysteresis and large abrupt changes in flow topology in one regime; the other regime is predicted to exhibit monostable behavior with a continuous change of flow topology. The predictions are confirmed in fully nonlinear numerical simulations of the 2D Navier-Stokes equation. Finally, implications of this work are briefly considered.

Large-scale flows in the atmosphere and ocean are characterized by stable density stratification, rotation, and small vertical-to-horizontal aspect ratios. These characteristics render turbulence at large scales quasi-2D, leading to a cascade of energy to large scales—a feature that is in contrast to the cascade of energy to small scales in commonly occurring, smaller-scale, 3D turbulence. A consequence of the inverse cascade is the self-organization of turbulence to form long-lived (large-scale) coherent structures such as the Great Red Spot and other vortices and jets in the Jovian atmosphere, the subtropical jet-stream and other vortical structures that are associated with storm systems in the earth's atmosphere, and mesoscale eddies and alternating jets in the world's oceans. Mechanisms that explain the formation of such coherent structures are in part amenable to statistical mechanical considerations [1,2].

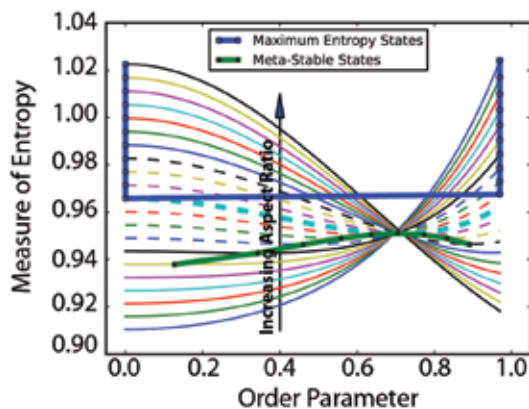


Fig. 1. Maximum entropy-based prediction for jet regime: The different smooth lines (without symbols) show the variation of a measure of entropy (y-axis) with order parameter (x-axis) at different aspect ratios. In this regime, the blue line connecting the filled blue circles (maximum entropy states) jumps abruptly as the aspect ratio is varied. The presence of meta-stable states over an intermediate range of aspect ratios indicates bistable behavior and hysteresis.

In this work [3], in the idealized setting of 2D turbulence, we consider the effect of domain geometry on flow topology. In particular, we consider regimes of turbulence that are conducive to the formation of two prototypical self-organized structures—jets or vortices—and apply the maximum entropy principle of statistical mechanics to produce a theory of flow topology in the two cases as the domain aspect ratio is varied.

Figures 1 and 2 summarize the results of the theory. In these figures, a measure of entropy is plotted as a function of the order parameter at a given value of aspect ratio; this is repeated for various values of the aspect ratio (different smooth lines). The filled blue circles indicate the “maximum entropy state” for each value of the aspect ratio, whereas the stars indicate meta-stable states if present at that value of aspect ratio. The blue line connecting the filled blue circles is the locus of

maximum-entropy states as the aspect ratio is varied. In Fig. 1, we consider the regime of turbulence that leads to the jet state; in Fig. 2, the regime of turbulence leads to vortex dipoles. In the jet-regime (Fig. 1) the locus of maximum entropy states exhibits a discontinuous jump as a function of aspect ratio, whereas in the vortex-regime (Fig. 2) there are no discontinuous jumps—the maximum entropy states continually deform as a function of the aspect ratio. Furthermore, in the jet-regime, there are meta-stable states for a range of aspect ratios (green stars), allowing for bistable behavior and hysteresis over this range of aspect ratios.

Figures 3 and 4 show the numerical verification of the theoretical predictions of Figs. 1 and 2 in a fully nonlinear setting. In Figs. 3 and 4, a measure of the order parameter is plotted as a function of the aspect-ratio parameter, as obtained in fully nonlinear simulations of the 2D Navier-Stokes equations. Also indicated in the colored panels are the long-time, self-organized, quasi-stationary flow patterns that result in the two different regimes and how they deform as a function of the aspect ratio. In the jet-regime of Fig. 3, the discontinuous jump from flow in the x-direction to flow in the y-direction is evident, as is also the regime of bistability over a range of aspect ratios surrounding 1, all in accordance with the theoretical predictions of Fig. 1. On the other hand, the fully nonlinear simulations of the 2D Navier-Stokes equations in the vortex-regime confirm that, in this regime, the flow topology changes continuously as a function of the aspect ratio, again in accordance with the theoretical predictions of Fig. 2.

In future work, we will build upon the above results to develop an explanation of the low-frequency regime transitions that have been observed in the non-equilibrium setting of this problem [4,5]. Consequently, insights developed here should be useful in developing a better understanding of the phenomenon of low-frequency regime



transitions, which are a pervasive feature of the weather and climate systems. Familiar occurrences of this phenomenon—wherein extreme and abrupt qualitative changes occur, seemingly randomly, after very long periods of apparent stability—include blocking in the extra-tropical winter atmosphere, the bimodality of the Kuroshio extension system, the Dansgaard-Oeschger events, and the glacial-interglacial transitions.

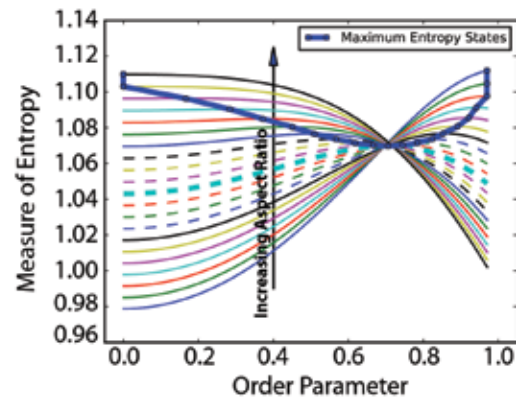


Fig. 2. Predictions for vortex regime: Flow topology evolves continuously as a function of aspect ratio. Consequently, there are neither abrupt transitions nor hysteretic behavior.

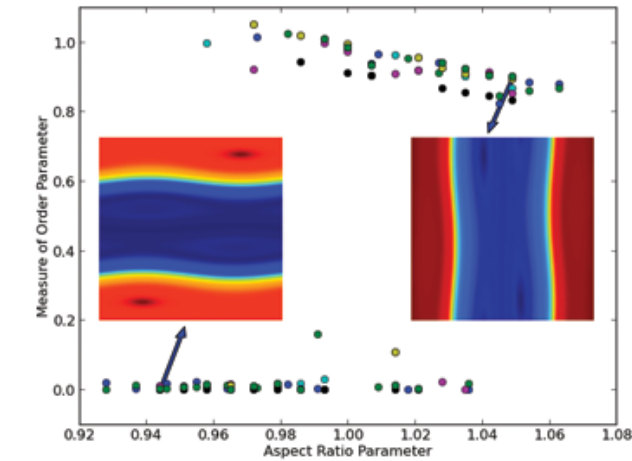


Fig. 3. Order parameter as a function of aspect ratio as obtained in fully nonlinear simulations of the 2D Navier-Stokes equations. Jet regime, abrupt transitions, bistable regimes, and hysteretic behavior are observed in accordance with the theoretical predictions of Fig. 1. Flow configurations on the two branches are shown in color insets.

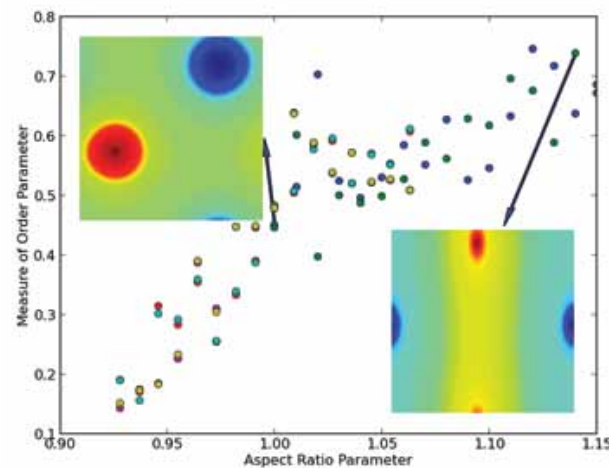


Fig. 4. Results of nonlinear simulations in the vortex regime. Continuous transformation of flow topology with geometry is observed in accordance with theoretical predictions of Fig. 2.

[1] Onsager, L., *Nuovo Cimento* **6**, 279 (1949).

[2] Kraichnan, R.H. and D. Montgomery, *Rep Progr Phys* **43**, 35 (1980).

[3] Loxley, P.N. and B.T. Nadiga, "Bistability and Hysteresis of Maximum-Entropy States in Decaying Two-Dimensional Turbulence," *Phys Fluids*, submitted (2012).

[4] Bouchet, F. and E. Simonnet, *Phys Rev Lett* **102**, 094504 (2009).

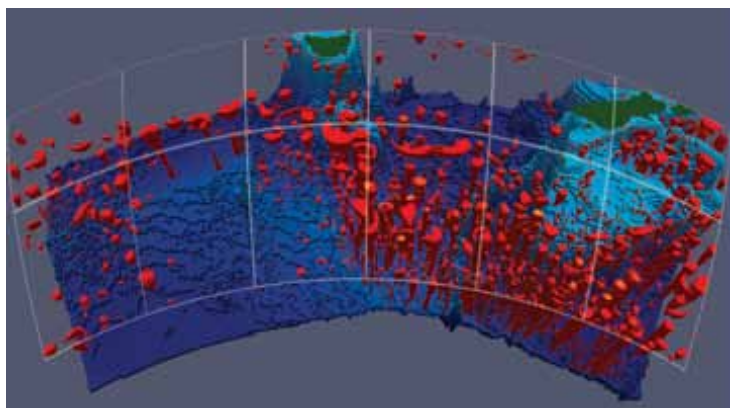
[5] Nadiga, B.T. and T.J. O'Kane, "Predictability of Low-Frequency Regime Transitions in the Barotropic Vorticity Equation Driven by Weak Stochastic Forcing," *Tellus Dyn Meteorol Oceanogr*, submitted (2012).

# A Three-Dimensional Eddy Census of a High-Resolution Global Ocean Simulation

**Mark R. Petersen, CCS-2;**  
**Sean J. Williams, CCS-3;**  
**Matthew W. Hecht, CCS-2;**  
**Mathew E. Maltrud, T-3;**  
**Bernd Hamann, Institute**  
**for Data Analysis and**  
**Visualization, University of**  
**California**

A 3D eddy census data set was obtained from a global ocean simulation with one-tenth degree resolution and a duration of seven years. Variables of interest include eddy diameter, thickness (vertical extent), minimum and maximum depth, location, rotational direction, lifetime, and translational speed. Distributions of these traits show a predominance of small, thin, short-lived, and slow eddies. Still, a significant number of eddies possess traits at the opposite extreme—thousands of eddies larger than 200 km in diameter appeared in daily data each year. We developed a new method of eddy identification that is well suited for global, 3D data sets and compares better with observations than previous methods. The eddy identification method has allowed us to locate and characterize 27 million eddies in the daily data and conduct detailed comparisons with surface observations.

**H**ow deep are ocean eddies? Do they look more like thin disks or tall columns? Do eddies with large surface extents tend to be deeper as well? How many eddies are completely hidden below the surface? These questions are difficult to answer with current observational data.



*Fig. 1. Okubo-Weiss field in the Southern Ocean to the south of Tasmania and New Zealand, showing 120E-180E and 45S-55S. The Antarctic Circumpolar Current is the region with the largest number of eddies and the deepest eddies in the world. Many of these eddies extend to the full depth of the ocean, others are strictly surface features, and some are completely submerged. The R<sup>2</sup> method is more discriminating and will eliminate many of the more spurious features seen here. Depth is exaggerated by a factor of 50.*

Detailed eddy characteristics are available from satellite altimetry, but provide no information about depth. Shipboard observations provide some hints, but are limited to 2D sections and are often shallow in depth. Ocean floats are an important tool for collecting subsurface data and have begun to fill in gaps in recent years, but they provide only a few profiles for each eddy.

Numerical simulations of the ocean provide full 3D velocity and tracer fields that lend themselves to automated eddy census and tracking algorithms. The purpose of this work is to characterize eddies of the global ocean—in particular, properties involving depth that are somewhat sparse in observational studies. To our knowledge, this is the first such eddy census of a global ocean simulation. Past work on vertical eddy structure is limited to regional domains on continental shelves.

The characterization of ocean eddies is the first step towards understanding their effects in the transport of heat, salt, chemical species, and organisms. Observational studies have shown that

discrete eddies can have a large impact on biological productivity—picophytoplankton biomass in individual eddies can be 30% higher than that of the surrounding waters.

The eddy census was conducted using velocity data from seven years of a high-resolution (one-tenth degree) simulation of POP (Parallel Ocean Program), developed and maintained by the Climate, Ocean, and Sea Ice team (COSIM) at LANL [1,2]. Current methods of eddy identification often use the Okubo-Weiss parameter, a measure of strain versus vorticity (Fig. 1). This method is unsuitable for global simulations because it detects many extraneous features and requires an arbitrary threshold that varies across the globe. A new method of eddy identification was developed, named R<sup>2</sup>, that judges the fitness of a vortex based on similarity of characteristics with an idealized Gaussian vortex [3-5]. Eddies identified by the R<sup>2</sup> method are more realistic than those chosen by Okubo-Weiss when compared to observational data due to the retention of fewer spurious features.

Using the R<sup>2</sup> method, an eddy census database consisting of 27 million eddies was created from seven years of daily data [6]. This is an unprecedented amount of data on eddies in the global ocean. Statistics of eddy characteristics were analyzed and compared to observations, when available, and a tracking algorithm was created to find eddy lifetimes, and limit the study to long-lived eddies of at least four weeks [6].

Eddy diameter is a strong function of latitude, with smaller eddies near the poles and larger ones near the equator (Fig. 2). This is expected since the first baroclinic Rossby radius varies strongly with latitude and is also consistent with satellite observations. The Antarctic Circumpolar Current contains the thickest and highest density of eddies (Fig. 2).

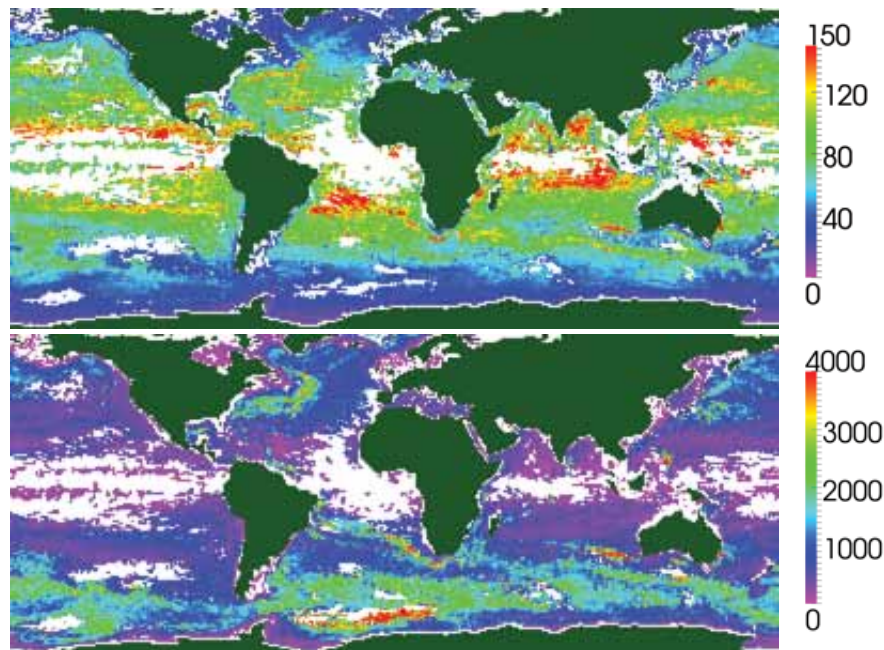


Fig. 2. Eddy statistics from seven years of a POP ocean simulation using the  $R^2$  eddy detection method, a minimum lifetime of four weeks, and collated in one-degree bins: diameter, km (top); thickness, m (bottom). The Antarctic Circumpolar Current stands out as the region with the most eddies and the deepest eddies. Deep eddies also appear in the Gulf Stream, Kuroshio Current, and Agulhas Ring pathway. White areas are one-degree cells where no eddies were detected over the seven-year census.

Thick eddies are also common in the Gulf Stream, Kuroshio Current, and Agulhas ring pathway.

A significant number of eddies penetrate deep into the ocean: a third of the eddies in this simulation are at least 1000 meters tall. Of eddies with a minimum four-week lifetime, the majority (97%) extend all of the way to the surface (Fig. 3). Although not all of these surface-expressed eddies located by the  $R^2$  method are clearly reflected in the surface height, it is very likely that satellite altimetry-based assessments of eddy size, spatial distribution, and lifetime are reasonably comprehensive as estimates of eddy characteristics. The remaining eddies that do not reach the surface are distributed over the full depth of the ocean, with thousands deeper than 3000 meters. Larger-diameter eddies are likely to be thicker, longer-lived, and faster than smaller-diameter eddies. Correlations between thickness and lifetime or thickness and speed are weak, except that very thin eddies are fast and shorter-lived.

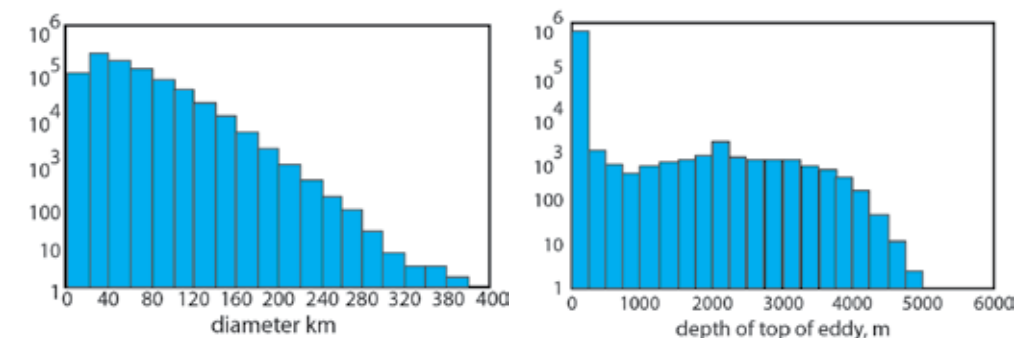


Fig.3. Distribution of eddies detected by diameter (left) and depth of the top of the eddy (right), using the  $R^2$  method and minimum lifetime of four weeks. This is the population of eddies recorded in daily averages, and vertical axes display number of eddies per year. The majority of eddies are small and thin, but there are still thousands of eddies with diameters greater than 200 km, and tens of thousands with thicknesses of 4000–5000 meters. The great majority extend to the surface, but tens of thousands exist below the surface.

We can confidently conclude that eddies are a common phenomenon in the deep ocean, albeit in smaller numbers than thin eddies near the surface. Observational studies of eddy transport of heat and nutrients have been confined to the upper ocean for practical reasons. The next step in the analysis of this simulation is to quantify the impact of discrete eddies on the transport of tracers throughout the globe. Indeed, high-resolution ocean model output provides a unique opportunity to compute detailed statistics where observations are sparse.

- [1] Maltrud, M.E. et al., *Environ Fluid Mech* **10**, 275 (2010).
- [2] Maltrud, M.E. and J.L. McClean, *Ocean Model* **8**, 31 (2005).
- [3] Williams, S. et al., *Comput Graph Forum* **30**, 991 (2011).
- [4] Williams, S. et al., *IEEE Trans Visual Comput Graph* **17**, 2088 (2011).
- [5] Williams, S. et al., *Comput Graph Forum* **31**, 1125 (2012).
- [6] Petersen, M. et al., "A three dimensional eddy census of a high-resolution global ocean simulation." *J Geophys Res*, submitted (2013).



# A Multi-Resolution Approach to Global Ocean Modeling

Mark R. Petersen, CCS-2;  
Todd Ringler,  
Doug Jacobsen,  
Mathew E. Maltrud,  
Philip W. Jones, T-3

The Model for Prediction Across Scales-Ocean (MPAS-Ocean) is a new global ocean model capable of using enhanced resolution in selected regions of the ocean domain. Two simulations are presented to evaluate the model: a uniform high-resolution (15 km) mesh; and a second grid with similarly high resolution (15 km) in the North Atlantic, but coarse resolution elsewhere. Comparisons with observational data show that currents and mesoscale eddy activity are well simulated in both structure and amplitude. Simulations using the variable-resolution second grid are essentially identical to the uniform case within the North Atlantic region. The overall conclusion is that this ocean model is a viable candidate for multi-resolution simulations of the global ocean system on climate-change time scales.

Over the relatively short history of global ocean modeling, the approach has been almost entirely based in structured meshes, conforming quadrilaterals, and a desire to obtain quasi-uniform resolution. This approach has the advantage that numerical schemes and data analysis are straightforward. The disadvantage is that resolution must be nearly uniform across the globe, and doubling resolution requires an additional factor of ten in computational resources.

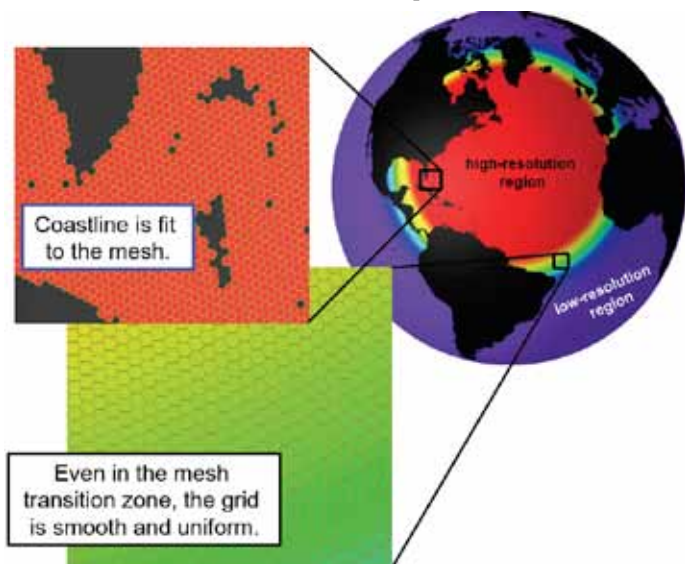


Fig. 1. The x5.NA.15km mesh includes a high-resolution North Atlantic region with 15-km grid cells, 75-km grid cells on the rest of the globe, and a smooth transition zone between them.

Here we present a new ocean model, MPAS-Ocean, developed by the Climate, Ocean, and Sea Ice Modeling (COSIM) team at LANL. Horizontal meshes are unstructured so that high-resolution regions may be embedded within a lower-resolution global mesh. This means high-resolution regional simulations may be performed at a lower computational cost, but the global simulation still supplies realistic currents and water properties from the low-resolution regions. The global meshes, created using Spherical Centroidal Voronoi Tessellations [1,2], vary smoothly from low to high-resolution regions (Fig. 1). Numerical algorithms specifically designed for

these grids guarantee that mass, tracers, potential vorticity, and energy are conserved [3,4].

In order to showcase the regional modeling capability of MPAS-Ocean, we present two simulations [5]: (1) x1-15km, a global quasi-uniform

mesh where gridcell width is about 15 km; and (2) x5-NA-15km, a variable-resolution mesh with 15-km gridcells in the North Atlantic and 75 km elsewhere (Fig. 1). A snapshot of kinetic energy for x1-15km (Fig. 2) shows many realistic features: Agulhas rings in the South Atlantic, a strong equatorial undercurrent in the Pacific, vortex ring shedding in the region off west Australia, tropical instability waves, and an Atlantic equatorial undercurrent that is fed via retroflexion of the north Brazil current, which periodically sheds coastally trapped rings that propagate into the Caribbean.

A number of quantitative comparisons between observations and the simulations show very good agreement [5]. Because the ocean is a chaotic system, comparisons must be made of decadal averaged climatological variables. Global observations of sea surface height (SSH) mean and variability are available from satellite altimetry and satellite gravity-recovery data sets. All the large-scale features of global ocean circulation are present in the simulations, with similar structure and amplitudes (Fig. 3).

Volume transport across specific cross-sections of the ocean have been observed by shipboard cruises. Decadal averages of these same sections in the simulations show similar ranges as the observations, and some key through-flows, like the Drake Passage, agree within 10% [5]. Cross-sections of the equatorial undercurrent in the tropical Pacific and the Deep Western Boundary Current east of the Bahamas show very good agreement between time-averaged observations and simulations.

An important component of this study is the comparison between the global uniform-resolution and regional variable-resolution simulations. For the North Atlantic region, mean SSH and SSH variability are nearly identical (Fig. 3), and transports through Caribbean passages are within

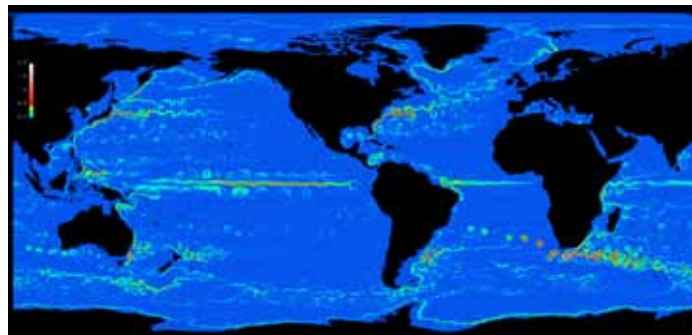


Fig. 2. A snapshot of kinetic energy at a depth of 100 m for the x1-15km simulation. The color scale saturates at red where instantaneous velocities reach 1.0 m/s.

10–20% [5]. This shows that specific regions of the global ocean system can be accurately simulated with local mesh refinement.

Efficiency and scaling on high-performance computers

with thousands of processors is an important design consideration for MPAS-Ocean. The x1-15km achieves two simulated years per wall-clock day on 3000 processors on Mustang, a large cluster at LANL. This is the same speed as high-resolution 0.1-degree POP simulations, a long-standing structured global ocean model. The MPAS-Ocean code is designed for scalability and porting to advanced heterogeneous architectures.

The MPAS-Ocean development team has adopted software engineering practices to ensure high-quality code. All projects begin with a requirements and design document, which is reviewed by the team before coding begins. Each new code branch is reviewed and tested by another team member before merging back to the trunk. In addition, a set of standard, automated test cases have been developed to ensure that the code meets design requirements.

One of the challenges of variable-resolution climate modeling is to ensure that turbulence closures and parameterizations are well implemented across scales. COSIM scientists are developing scale-aware parameterizations [6] that will benefit all modeling efforts with methods that reduce the need for tuning parameters for a particular grid scale.

MPAS-Ocean is one component within the MPAS framework of climate models that is developed in cooperation between LANL and the National Center for Atmospheric Research

(NCAR). Other MPAS components include land-ice and sea-ice, both under development by the COSIM team at LANL, and atmospheric components created by NCAR. In the coming years, these components will be coupled within the Climate Earth System Model (CESM) to explore research questions surrounding anthropogenic climate change. For the first time, variable-resolution-coupled models of the atmosphere, ocean, sea-ice, and land-ice will be available to study the impacts of climate change on specific regions at very high resolution.

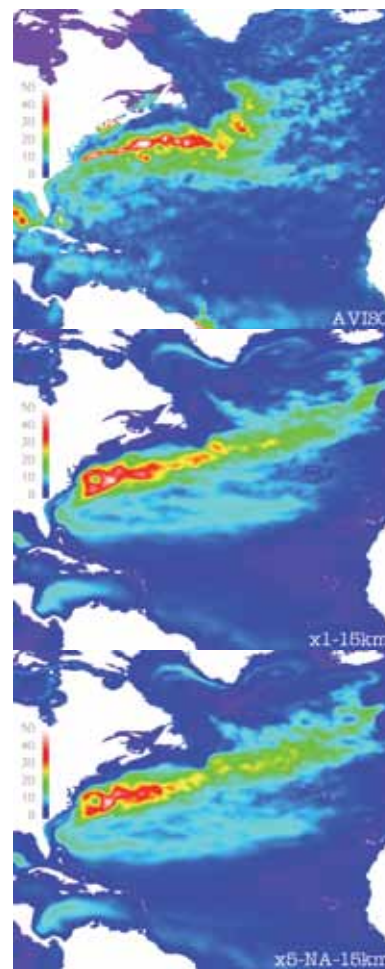


Fig. 3. Sea surface height variability in the North Atlantic from observations (top), the globally uniform mesh (x1-15km, middle), and the North Atlantic regional simulation (x1-NA-15km, bottom). All panels use the same color scheme ranging from 0 cm to 50 cm.

- [1] Ringler, T. et al., *Ocean Dynam* **58**, 475 (2008).
- [2] Ringler, T. et al., *Mon Weather Rev* **139**, 3348 (2011).
- [3] Thuburn, J. et al., *J Comput Phys* **228**, 8321 (2009).
- [4] Ringler, T. et al., *J Comput Phys* **229**, 3065 (2010).
- [5] Ringler, T. et al., “A Multi-Resolution Approach to Global Ocean Modeling.” *Ocean Model*, in press (2013).
- [6] Chen, Q. and T. Ringler, *ADTSC Science Highlights*, 36, (2013)

# Two Modes of Sea-Ice Gravity Drainage

Adrian K. Turner,  
Elizabeth C. Hunke, T-3;  
Cecilia M. Bitz, University of  
Washington

Processes that change the vertical salinity profile of sea ice have a significant impact on sea-ice properties and biogeochemistry. One of the most important processes affecting sea-ice salinity is gravity drainage where cold, dense brine within newly formed sea ice drains out and is replaced by seawater, resulting in a significant desalination of the sea ice. Using time-resolved bulk salinity and temperature observations of forming sea ice, we show that gravity drainage occurs as two distinct modes: (1) rapid drainage at the base of the ice and (2) slow drainage occurring more deeply in the ice. We have developed a parameterization for gravity drainage that includes these two modes and is suitable for inclusion in a global climate model. The parameterization is included in an all-new thermodynamic component, based on mushy layer theory, for the LANL sea ice model (CICE).

Sea ice, the frozen surface of high latitude seas, is not entirely fresh—it is composed of a network of brine pockets surrounded by fresh ice. The brine is not fixed in the sea ice—brine inclusions expand and contract as the temperature of the ice changes and the fresh ice surrounding the brine network melts and refreezes. Brine pockets within the sea ice are connected together as a porous medium, and a number of processes move the brine around and change the properties of the sea ice. The brine also harbors a rich variety of living organisms whose existence in the ice is enabled by the transport of nutrients into the ice with seawater through the brine network.

The vertical salinity profile of sea ice changes shape during its first year, from a “C” shape with higher salinities at the top and the bottom than in the interior, to a profile characteristic of multiyear ice, with much fresher ice near the top surface and less salt content overall. The salinity structure of the ice directly affects heat conduction and melt/freezing rates and strength and has indirect effects through the effect of the sea-ice biology (changing albedo for example). As the seasonal ice fraction of the Arctic ice pack increases [1], the representation of the early evolution of the sea-ice salinity profile becomes more important in models. In the Southern Hemisphere, where the pack ice is already largely seasonal, such a representation will improve simulations of global and regional climate.

One of the most important processes that changes the salinity profile of the ice is gravity drainage. When sea ice forms, the upper layers are colder (and so contain denser brine) than deeper layers. This unstable density profile results in the brine draining out of the ice to be replaced by fresher seawater, resulting in a significant desalination of the ice.

Until now, the salinity profile in CICE was fixed in a form appropriate for multiyear ice that has already largely drained. We present a new 1D parameterization of gravity drainage implemented in an all-new thermodynamic component of CICE, based on mushy layer theory [2]. We solve a set of coupled, nonlinear equations for sea-ice temperature (enthalpy) and salinity using an implicit Jacobian-free Newton-Krylov method.

Time-resolved data [3] reveal two distinct modes of sea-ice drainage: (1) rapid drainage in a narrow region at the base of the ice and (2) slow drainage occurring more deeply in the ice. In this paper we present a prognostic salinity parameterization suitable for inclusion in a global climate model that reproduces both drainage modes apparent in the observed data. The rapid mode of drainage is modeled with an advective operator assuming upflow in the mush and downflow through evacuated channels. The slow desalination is represented as a simple relaxation of bulk salinity to a value based on a critical porosity for sea-ice permeability. We find that these parameterizations can adequately reproduce observational data from laboratory experiments and field measurements.

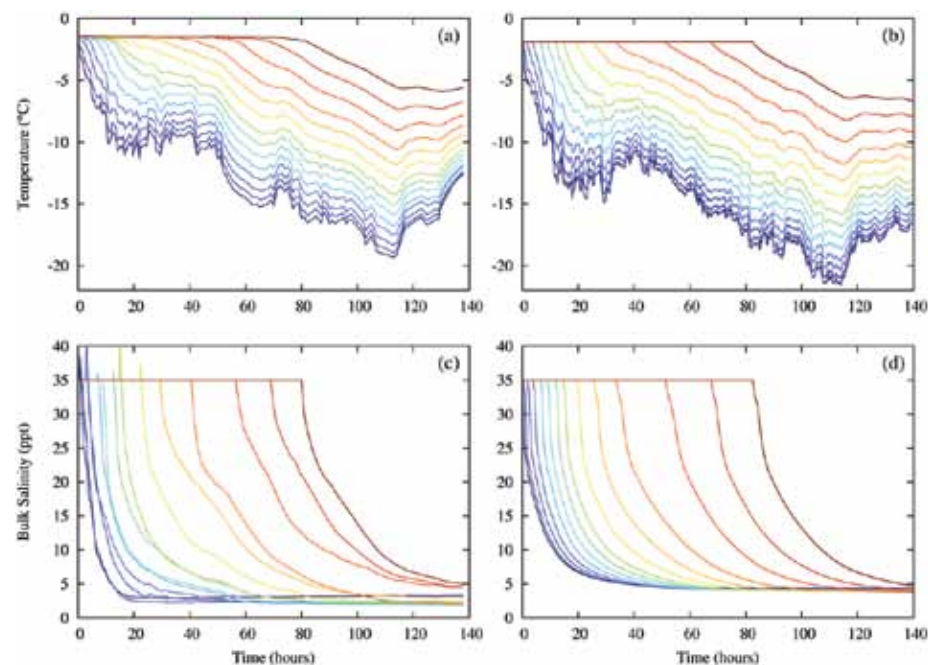
These observations make use of an instrument capable of measuring the temporal evolution of bulk salinity at various depths within sea ice as it forms [4]. The instrument consists of 14 pairs of platinum wires placed at various depths within the ice (6, 11, 18.5, 26, 36, 46, 56, 66, 81, 96, 111, 131, 151, and 171 mm below the upper ice surface). The electrical impedance is measured between the pairs and is used to infer the liquid fraction of the ice. The instrument also has a series of thermistors that record temperature at the same depths as the wire pairs that measurements are converted to brine salinity using a liquidus relation. With the measured liquid fraction and brine salinity, bulk salinity may



be inferred. Notz and Worster [3] used this instrument in Adventfjorden in Spitsbergen during the winter of 2005 to measure sea ice temperature and bulk salinity.

Figure 1 compares the observed and modeled temporal evolution of temperature and bulk salinity at each of the instrumental depths—dark blue colors correspond to sensors near the top surface of the ice. As the ice grows, deeper sensors are gradually incorporated. The two stages of gravity drainage are evident in the lower panels of Fig. 6, showing rapid initial drainage that tapers off into the slow drainage mode at each sensor level. The temperature and salinity fields are reasonably reproduced by the model.

The rapid drainage parameterization presented here is based on an advective operator, unlike other gravity-drainage formulations that have generally used a diffusive parameterization. This more closely models the physical reality of gravity drainage that is thought to consist of an upward Darcy flow in the mush and downward flow through narrow channel features. Diffusion implies mixing of downward and upward flows at each level, something that is not thought to happen in gravity drainage. The parameterization considers explicitly the pressure differences in the mush and channels that drive the flow, and it successfully captures the observed sea ice salinity evolution in a manner suitable for global climate modeling.



*Fig. 1. Temperature (degrees C) versus time (hours) for (a) the individual wire pairs in field observations [3], and (b) the corresponding model simulation with 100 layers. Bulk salinity (ppt) versus time (hours) for (c) the individual wire pairs in field observations [3], and (d) the corresponding model simulation with 100 layers. Model data are interpolated to the same depths as the wire pairs in the experiment.*

[1] Comiso, J.C., *J Clim* **25**, 1176; doi:10.1175/JCLI-D-11-00113.1 (2012).

[2] Turner, A. K. et al., "Two Modes of Sea-Ice Gravity Drainage," LA-UR 12-23270; *J Geophys Res*, submitted (2012).

[3] Notz, D. and M.G. Worster, *J Geophys Res* **113**, C03001; doi:10.1029/2007JC004333 (2008).

[4] Notz, D., "Thermodynamic and Fluid-Dynamical Processes in Sea Ice," Ph.D. thesis, University of Cambridge, U. K. (2005).

# Modal Variability in the Bellingshausen Basin

Wilbert Weijer, CCS-2

Satellite observations revealed an exceptional circulation pattern in the Bellingshausen Basin in late 2009. Here we study that event using a one-layer ocean model forced by observed winds. We find that the event can be explained as a circulation pattern trapped by the local bathymetry, especially the Bellingshausen Basin and East Pacific Rise. An analysis of the energetics suggests that the mode loses most of its energy at only a few areas, like the Eltanin Fracture Zone and the crest of the East Pacific Rise. These areas might be “hot spots” for abyssal mixing.

During the fourth quarter of 2009 (Q4 2009), exceptionally high sea levels were detected in the southeastern South Pacific Ocean, in an area known as the Bellingshausen Basin [1]. The signal had an amplitude of about 8 cm and persisted for several months. It was found in radar altimetry, as well as in bottom pressure estimates from the Gravity Recovery and Climate Experiment (GRACE) satellite mission.

The Bellingshausen Basin (Fig. 1), together with Mornington and Amundsen Abyssal Plains, form a deep abyssal basin in the southeastern South Pacific Ocean. High sea-surface height (SSH) variability has been

detected there before [2], and was interpreted as a circulation pattern trapped by the local bathymetry [3]. In particular, ocean circulation tends to be steered by bathymetry—currents do not like to flow across isobaths since the resulting stretching or squeezing of the water column requires a change in a quantity called vorticity (the continuum-mechanical equivalent of angular momentum). So water preferentially flows along contours of so-called potential vorticity, or  $f/H$  (where  $f$  is twice the local rotation rate of the Earth and  $H$  is ocean depth). These contours are strongly warped by the bathymetric features of the southeastern South Pacific (contours in Fig. 1) and, in fact, many contours are closed or almost closed. This gives rise to the possibility of modal circulation—circulation along such (almost)

closed contours decays very slowly, retaining their large-scale spatial structure for days or weeks. Depending on whether the circulation is clockwise or anticlockwise, it is associated with low or high sea level in the center. And it is just such a signal that was recorded in Q4 2009.

Here we study the excitation and energetics of this mode in a so-called shallow water model. This model treats the ocean as a single layer of constant density, which is appropriate for the intraseasonal variability that we are interested in [4]. We force the model with wind stress observations from 2008 through 2010. The model displays substantial skill in reproducing the high sea level in late 2009. Figure 2 shows that the SSH in Q4 2009 was significantly positive over the Bellingshausen Basin and East Pacific Rise, and that the anomaly appears to be trapped by the distribution of  $f/H$ . The amplitude of the SSH anomalies seems somewhat underestimated (reaching 2 cm instead of 8 cm); however, the Q4 2009 event was clearly anomalous in its amplitude and persistence compared to the rest of the time series. A weaker event is visible during the third quarter of 2008 (Q3 2008).

What we are mainly concerned with is what the excitation of the mode means for the energy balance of the ocean. Wind forcing is the main source of energy for the ocean. The question is where this energy ends up. In recent years it has become clear that interaction of the ocean circulation with bathymetry may lead to the generation of small-scale turbulence, which can mix water masses upwards through the ocean column. Thus, it is believed that abyssal mixing plays an important role in the global thermohaline circulation. Figure 3 shows the energetics of the mode, compared to the non-modal circulation. Clearly, both the modal (solid lines) and non-modal (dashed) circulation receive most of their kinetic energy from the wind forcing (blue), and lose a lot through friction (black). However, there is a significant transfer of energy from the modal to the non-modal circulation through the work done by pressure forces (red). This indicates that the mode is decelerated (loses energy) through the emission of planetary waves wherever it encounters discontinuities in the contours of  $f/H$ . Additional analysis shows that such discontinuities are encountered in only a few places, in particular in the Eltanin Fracture Zone and on top of the East Pacific Rise. So in these

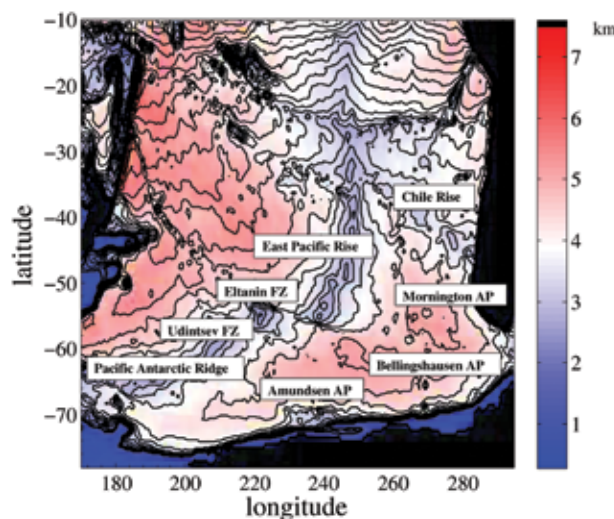


Fig. 1. Bathymetry of the South Pacific with some relevant geographic features indicated (FZ: Fracture Zone; AP: Abyssal Plain). Contours are isolines of potential vorticity,  $f/H$ .

locations the mode loses most of its energy and thus may be considered hot spots for abyssal mixing.

An interesting observation is that the wind input into the mode was apparently stronger during Q3 2008 than Q4 2009. Although Fig. 3 does show a persistent anomalous event during that period as well, it was not as strong as during Q4 2009, and is absent in the observations. The reason for this discrepancy is not clear.

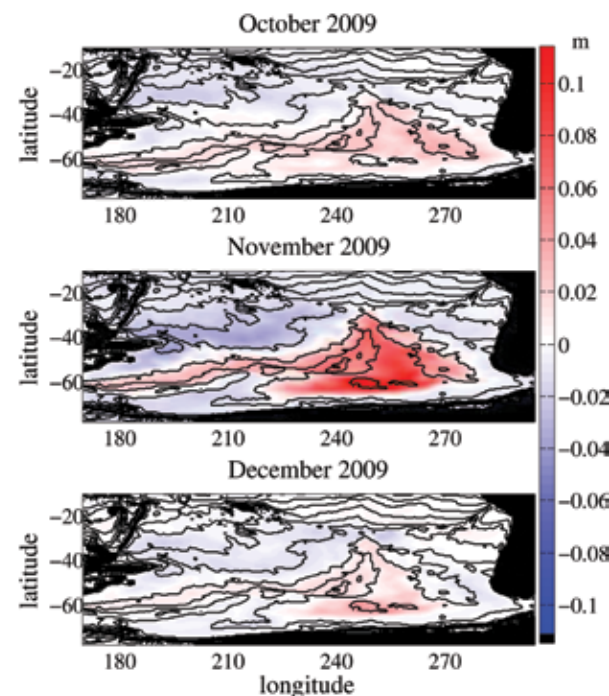


Fig. 2. Average SSH (m) for the months of October, November, and December 2009.

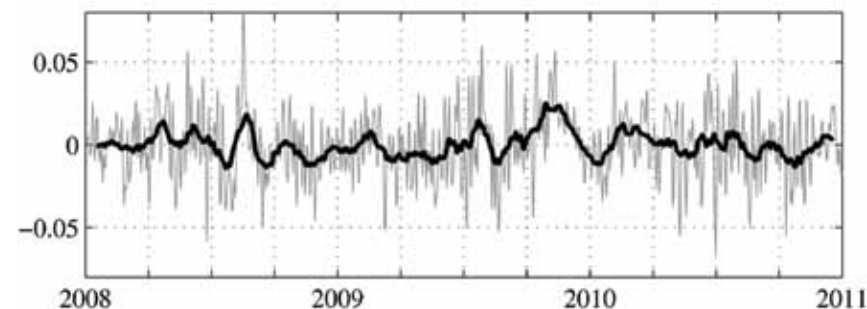


Fig. 3. SSH anomalies averaged over the region 90°W–140°W, 55°S–35°S.

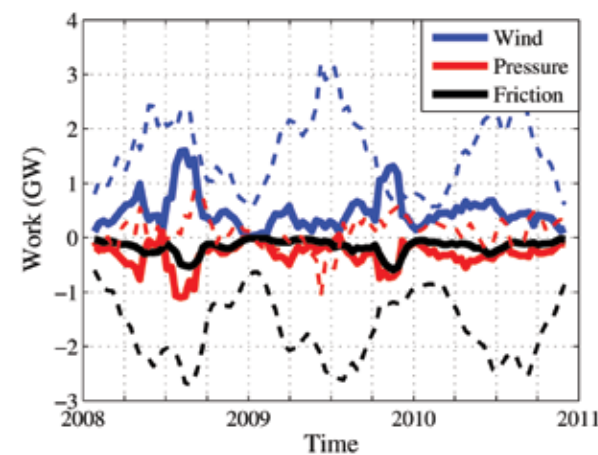


Fig. 4. Energy balance of the modal (solid lines) and non-modal, residual (dashed lines) circulation. Time series are 50-day moving averages.

- [1] Boening, C.T. et al., *Geophys Res Lett* **38**, L04602; doi:10.1029/2010GL046013 (2011).
- [2] Fu, L.L. and R.D. Smith, *Bull Am Meteorol Soc* **77**, 2625 (1996).
- [3] Webb, D.J., and B.A. de Cuevas, *J Phys Oceanogr* **33**, 1044 (2003).
- [4] Weijer, W., "Modal Variability in the Southeast Pacific Basin: Energetics of the 2009 Event." doi: 10.1016/j.dsr2.2012.10.002 *Deep Sea Research Part II*, in press (2012).



# Unified, Unstructured Grids for Climate Modeling

Geoffrey A. Womeldorff, T-3;  
Janet Peterson,  
Max Gunzburger, Florida  
State University;  
Todd Ringler, T-3

We have developed novel methods to generate unified, unstructured grids over the sphere with the goal of rendering the need to create separate grids for different applications obsolete and ushering in easier coupling of models while reducing coupling error. We create our spherical centroidal Voronoi tessellations and simultaneously their Delaunay duals with an iterative procedure. The most salient feature of our procedure is a complex density function that is able to take into account an arbitrary number of density proxies can represent physical measurements, model data, and/or error estimates. Thus, we can affect local refinement for different applications in different regions by way of multiple data sources. In addition, we take into account a representation of the shoreline in grid creation, allowing the triangulation to adapt itself to that representation to ensure that different applications will have appropriate interface regions.

The grids used to solve physical equations over the sphere, such as atmosphere and ocean models, have changed over the years and, moreover, the discretizations popular at a time inform the grid, and vice versa. A grid of latitude and longitude is the oldest and most widely used, marrying very naturally to both finite difference and finite volume methods. Over the last twenty years, in addition to seeing an increase in theory [1] and methodology to create a Voronoi diagram on the sphere, we have seen developments in grid generation and analysis of convergence of iterative methods and their acceleration. More recently, the first general circulation model was proposed over a Voronoi grid, and work continues in climate modeling utilizing Voronoi meshes and their duals [2]. In short, there exists a rich history of use of Voronoi (geodesic/hexagonal) grids in ocean and atmosphere modeling.

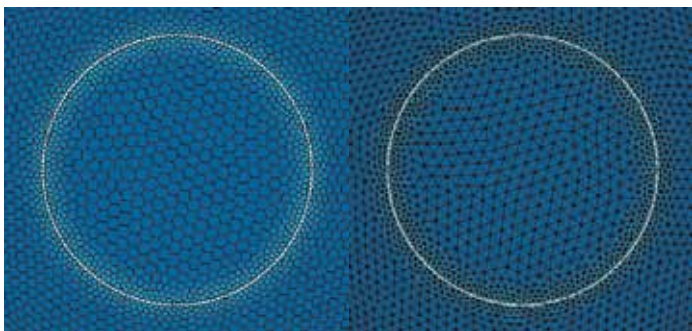


Fig. 1. A spherical centroidal Voronoi mesh and its Delaunay dual. The Voronoi cells straddle the interface while the triangle edges define it. Additionally, one can easily discern the increased density of cells around the interface.

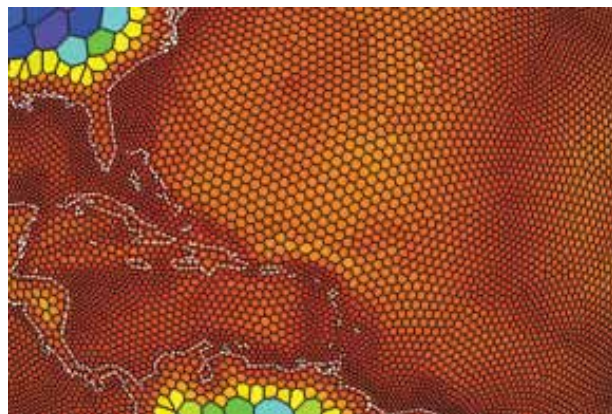
Until recently, however, Voronoi-type unstructured grids came with certain limitations. One of these is global refinement in discrete steps, with sizes related to repeated dissection. Additionally, without both local and global refinement, there are issues such as interpolation error and artifacting, related to coupling multiple equations such as ocean and sea ice.

To ameliorate the solution of the issues of coupling multiple equations over the sphere, we integrate a database of shoreline points into our grid generation process to demarcate various regions of the earth. Once we have sectioned off areas of the globe for different applications, we then add the ability to refine different sections based on arbitrary density proxies.

Methods exist that allow for fitting a triangulation to a boundary, but methods did not exist that allow for the use of multiple density proxies to be used to create unified, global grids suited to more than one application simultaneously. We utilize rich Voronoi grids that allow both global and local refinement in complementary ways. We create our grids using methods that allow for an arbitrary number of grid cells to be used, effecting global refinement to any degree desired, adding the ability to locally refine around shorelines, and to refine arbitrarily based on user input to better capture features of interest.

A spherical centroidal Voronoi tessellation (SCVT) is a special type of discretization of the sphere, defined by a set of points, generators, and a density function. These tessellations are generated by an iterative process. This iterative process allows us to tailor the grid by setting the initial conditions and the density function, leading to a very flexible system of mesh creation. Denoting a tessellation as Voronoi, along with its generators, implies that each region in the tessellation is composed of the subset of the area that is closer to the region's generator than any other. This Voronoi property provides for regions which are very regularly shaped. Adding centroidal as a qualifier to a Voronoi tessellation requires that each generator be the mass center of its region, and it is this property that requires iterative mesh creation. We use a custom density function to control the relative size of each region, as the density function is necessary in computing the centroids. Most importantly, a smoothly varying density function elicits a smooth change in region size that is, in general, a desirable property for nonuniform grids to have.

It is well known that the Delaunay triangulation and the Voronoi tessellation are planar duals of each other, thus the circumcenters of the triangulation are the vertices of the Voronoi regions and the mass centers of the Voronoi regions are the vertices of the triangulation. This



*Fig. 2. A local view of a global SCVT, focused on the Florida coastline and the central Atlantic. The density is regulated by two disjoint proxies, bathymetry measurements in the ocean, and altimetry over the land. Note refinement along the shoreline interface.*

means that, in our iterative methods, we can choose to compute either mesh and derive the dual directly from our iterative results. We use a Lloyd's iteration [3] to generate our SCVTs, described as: (1) start with a set of initial generators, (2) form a Voronoi diagram from the generators, (3) calculate the centroid of each Voronoi region using the density function, and (4) check for

termination criteria, and, if not satisfied, return to step (2). In practice, we exploit the dual relationship between our meshes by forming a Delaunay triangulation and then producing the Voronoi connectivity from that by walking around the neighbors of each vertex in each triangle.

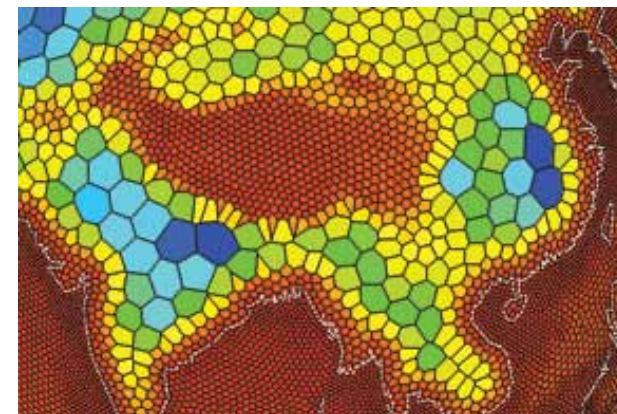
In order to create unified, unstructured grids that can service multiple equations at the same time, we must make use of the grid on both sides of the shoreline, and in order to size grid cells for multiple equations, we must be able to use an arbitrary number of density proxies. For example, we may wish to study the Gulf Stream, and need to refine the ocean grid in its area of influence around it, and simultaneously we may wish to refine for the Labrador Current. Another example might be to simultaneously refine a grid for an ocean model based on multiple criteria, such as sea surface temperature and potential vorticity. This is as simple as mapping the density proxy of interest, such as physical data, into the format used by our grid-generation software, and assigning it a bounding box to reside in. There are two main additions that we make to the iterative process that create our SCVTs: (1) a multiple proxy density function and (2) the concept of projecting points to the shoreline.

Shoreline (in general, boundary or interface) information is important when enabling multiple application codes

to communicate over the interface of disjoint regions that abut. We use this accurate interface region to remove the hurdle of having separate grids whose cells adjacent to the shoreline do not match. When cells on the interface do not align, complicated procedures are required to communicate data from one application to the other, often

increasing error or reducing convergence rates. We avoid these nonmatching interface region cells with our present technique. Additionally, we enforce a region of high density (i.e., small grid cells) in a band around shoreline interfaces. One could think of the band around the shoreline as being a safety net, for grid sizing, ensuring a safe maximum size around the shore, to provide for a smooth coupling of equations, and allowing the enforcement of grid size to safely relax inversely with the distance to the shore.

In the balance, we increase the quality of multi-domain, multi-physics applications. We do this by enabling local refinement around interface regions that adjoin, and in application-specific regions of interest, while gaining arbitrary global refinement—all of these features are made feasible through Voronoi methods. The most practical examples are climate related and on the sphere, but all of our ideas translate readily to more general spaces and dimensions.

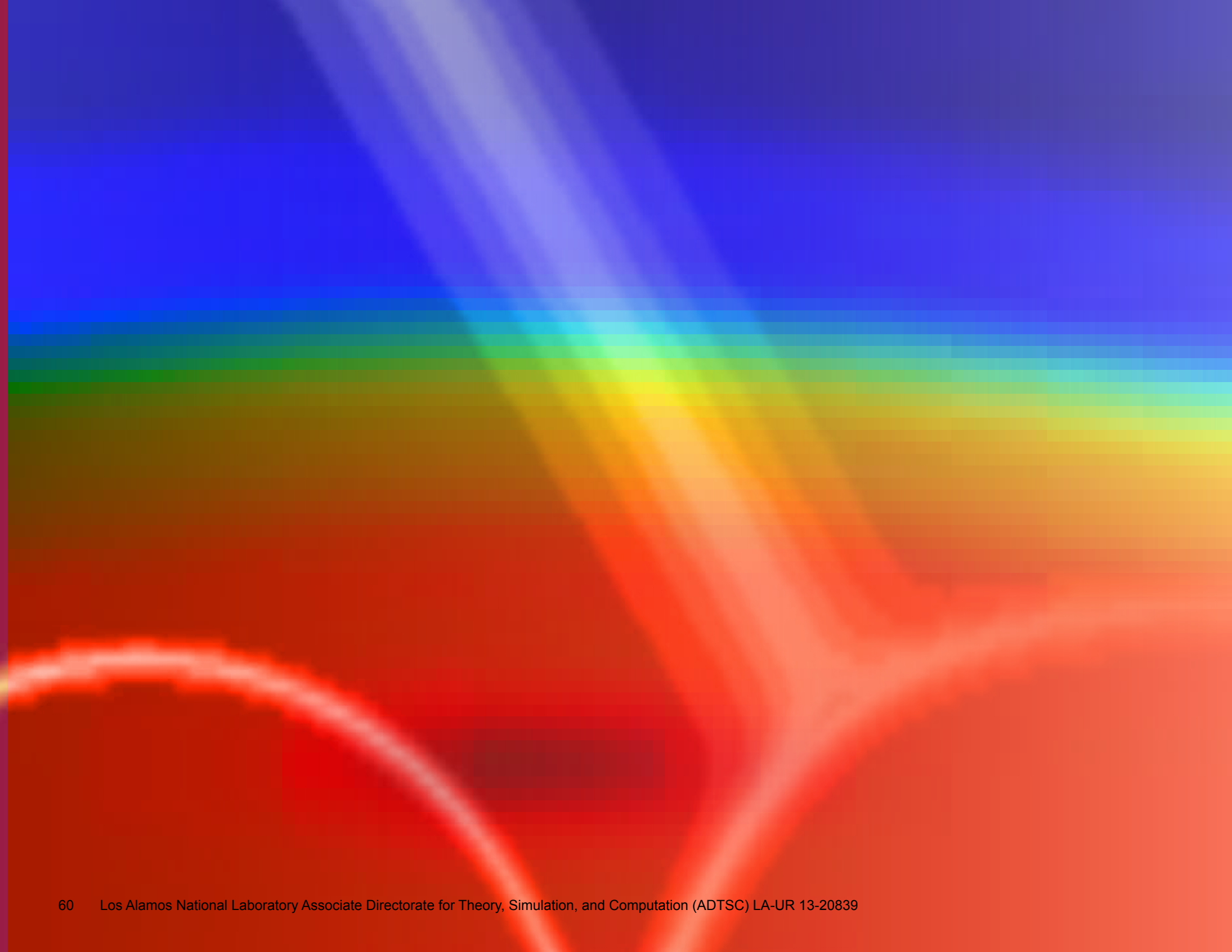


*Fig. 3. A local view of a global SCVT, focused on the Himalayan mountain region. The density is regulated by two disjoint proxies, bathymetry measurements in the ocean and altimetry over the land. Note refinement along the shoreline interface.*

[1] Du, Q. et. al., *SIAM Rev* **41**(4), 637 (1999).

[2] Ringler, T.D. et. al., *Mon Weather Rev* **139**(11), 3348 (2011).

[3] Lloyd, S.P., *IEEE Trans Inform Theor* **28**(2), 129 (1982).





# Information Science and Technology

**The Information Science and Technology field draws upon computer science, computational science, and mathematics to provide innovation in the areas of data processing and management and to derive actionable information from data. Information Science and Technology capabilities underlie most science endeavors at LANL.**

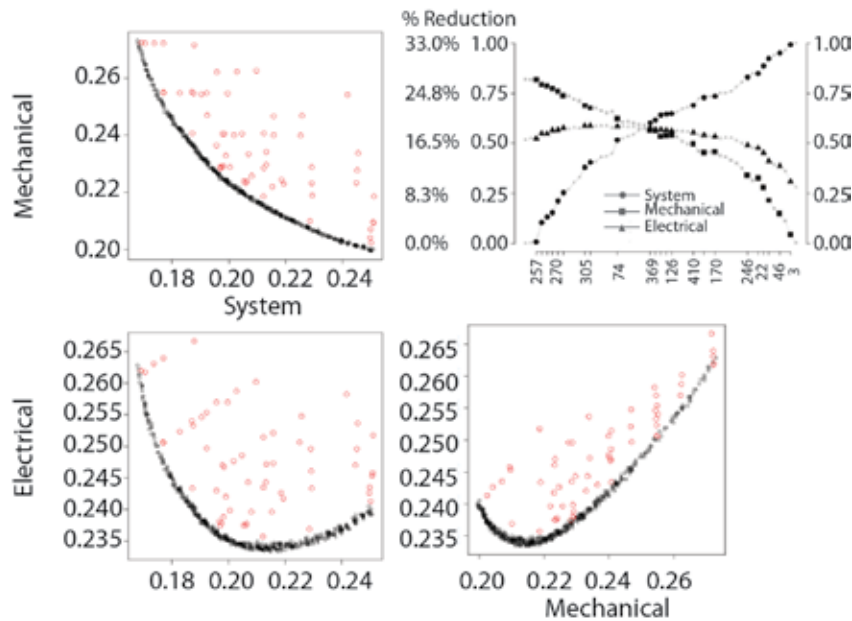
**This section includes contributions to the field of Information Science and Technology in the broad area of applied computer science. The 17 papers in this section range in breadth from quantum mechanics—applied to photosynthetic properties—to social interaction—applied to epidemiology and education. In between are several descriptions of tools and methods to be applied in scientific arenas relevant to LANL or against the various looming barriers in large-scale computation itself. Several papers address the movement of data in various ways; another paper addresses memory correction in an era of unprecedented-scale deployments; and three papers exploit the unique features of graphics processing units for future scientific gain.**

# Strategic Data Collection for Optimal Use of Resources Based on Multiple Criteria

Christine Anderson-Cook,  
CCS-6

The LANL Statistical Sciences Group, CCS-6, has an established record of research in experiment design and resource allocation. The goal of these efforts is to provide a quantitatively justified approach for determining what future data should be collected based on specific programmatic goals. The Pareto front approach allows decision-makers to establish specific objectives on which to focus for future data collection and, to identify superior choices to consider further, and then provide graphical summaries to compare alternatives with the goal of identifying a best choice. The approach allows flexibility for how the different objectives are weighted in the optimization, while allowing examination of the robustness across different prioritizations. This article describes a simplified example of how the methods can be used to maximally reduce uncertainty associated with system and subsystem reliability estimates.

Fig. 1. Pairwise scatterplot and trade-off plot of Pareto front for system and subsystem reliability estimation example.



As budgets become more tightly constrained, making thoughtful and justifiable decisions about what data to collect becomes increasingly important. In addition, there may be several types of data with different associated characteristics, costs, and precision from which to select. The Pareto front methodology adapted to experiment design and data collection [1] provides an approach for making quantitatively based decisions that can be tailored to the particular needs of a study. The process involves several steps. Step 1 requires that one or more top objectives for the study be identified and a quantitative metric determined that summarizes the performance based on each objective.

There is considerable flexibility to choose measures that represent different aspects of a good solution and are specific to the current study. Step 2 is an objective step where the Pareto front [2] is constructed (by selecting superior candidates either from a list of potential solutions or performing a formal optimization search). Potential solutions not on the front can be discarded from future consideration since they are strictly inferior to

at least one solution on the front and hence do not represent rational choices. Step 3 examines solutions on the front and allows examination of their performance on each of the criteria separately. In addition, the decision-maker can examine robustness to different desirability function weights [3], trade-offs between criteria, and overall performance relative to the best available for different weights [4]. Based on the specific needs of the study and understanding the alternatives, the decision-maker can select which data should be collected.

To illustrate the methodology, consider a series system composed of two subsystems (mechanical and electrical), with two and four components, respectively. There are nine different types of pass/fail data available (1 system + 2 subsystem + 6 component), each with different associated costs. (In previous years, some data were collected from each of these data types, and a Bayesian analysis [5,6] was used to estimate the reliability of the system and all of its parts). The primary goal of collecting new data from a fixed budget is to maximally improve the precision of the system reliability estimate. This can be quantified with the width of the 95% credible interval from the analysis. In addition to this objective, it is also desirable to improve the precision of the estimates for the two subsystems (as measured by the widths of their credible intervals).

Figure 1 shows a pairwise scatterplot of the Pareto front for these three objectives, along with a trade-off plot [1] to capture the relationship between objectives. In the scatterplots, the red circles represent proposed allocations of the resources, which were used to seed the search for best allocations using a genetic algorithm. The black circles represent solutions on the front, where ideal solutions would be located in the bottom left corner of each plot (minimizing the credible interval

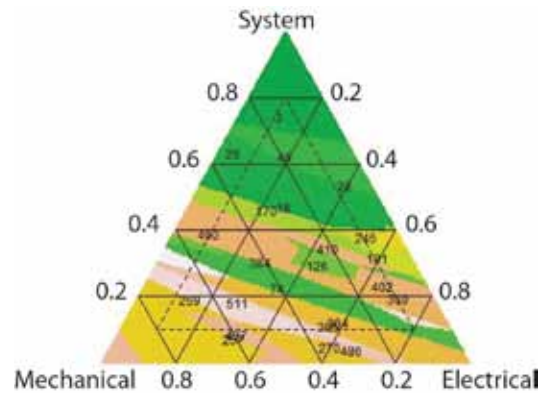


Fig. 2. Mixture plot of best allocations for different weight combinations of reductions for the system and subsystem credible interval widths.

maximizes improvement (approximately a 33% reduction) for the system, but performs relatively poorly for the two subsystems. Alternatively, in the middle of the plot, there are some solutions that achieve at least a 16% reduction in width for all three criteria.

Figure 2 shows a mixture plot of the allocations that are best for particular weight combinations of the criteria using the additive desirability function [3] scaling of objectives onto a common scale. The vertices correspond to placement of all the weight on a single criterion, the edge represents the combination of just two objectives with non-zero weights, and the interior corresponds to non-zero weights for all three criteria. Note how allocation 3 labeled and shaded in green at the top of Fig. 2, is the near-best for maximal reduction of the width of the system reliability credible interval,

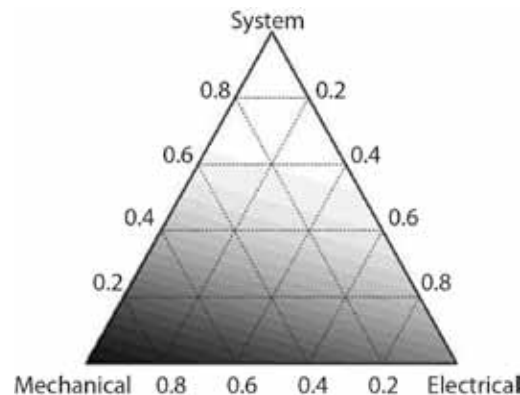


Fig. 3. Synthesized efficiency plot for a particular allocation to compare its performance against the best possible allocation for various weight combinations.

corresponds to a moderately large region near the System (top) vertex. Different alternatives are suggested as best depending on the decision-maker's valuation of the different contributions.

Finally, Fig. 3 shows the synthesized efficiency [4] of allocation 3 relative to the best design possible for each weight combination of the three criteria. This allows the decision-maker to determine the adequacy of its performance for the subsystem to objectives. For more details on the analysis and decision-making process for this example, see [7]. The suite of graphical methods helps describe the interrelationship between criteria and the possible performance across all of the objectives when considered together.

The advantages of the Pareto front approach include its flexibility for accommodating any possible quantitative measures for characterizing the goals of a data-collection study. This flexibility allows the decision-maker to solve the right problem. In other scenarios, we considered combining physical and first principles modeling data and optimizing based on good prediction, estimating tuning parameters and calibrating the data through a discrepancy function. In addition, by identifying the entire Pareto front of potential candidate solutions, the range of possible values for each criterion is available to calibrate comparisons. This method can be applied to a wide variety of data collection problems that allow leveraging current understanding of the problem and quantifying the anticipated benefit the additional data will provide in advance.

- [1] Lu, L. et al., *Technometrics* **53** 353 (2011).
- [2] Kasprzak, E.M. and K.E. Lewis, *Structural and Multidisciplinary Optimization* **22**, 208 (2001).
- [3] Derringer, G. and R. Suich, *J Qual Tech* **12**, 214 (1980).
- [4] Lu, L. and C.M. Anderson-Cook, *Qual Eng* **24** 404 (2012).
- [5] Anderson-Cook, C.M. et al., *J. of the Military Operations Research Society* **13** 5-18 (2008).
- [6] Chapman, J. et al., *Technometrics* **54** 410-424 (2012).
- [7] Lu, L. et al., "A Case Study on Selecting A Best Allocation of New Data for Improving the Estimation Precision of System and Sub-System Reliability using Pareto Fronts," LA-UR 12-21268 (2012).



# Non-Hermitian Approach for Modeling of Noise-Assisted Quantum Electron Transfer in Photosynthetic Complexes

Gennady P. Berman, T-4;  
Alan R. Bishop, PADSTE;  
Alexander I. Nesterov,  
Universidad de Guadalajara

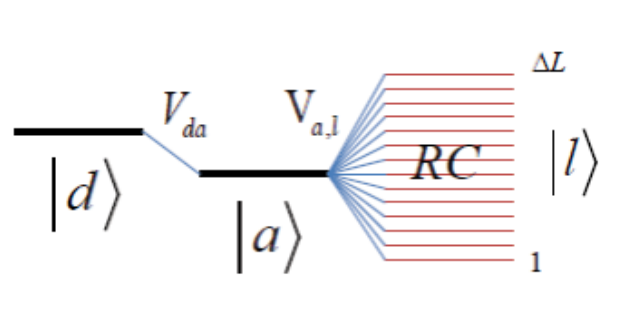
Nature has evolved photosynthetic organisms to be extremely complex bio-engines that capture visible light in their peripheral light-harvesting complexes (LHC) and transfer excited-state energy (as excitons) through the proximal LHC of photosystem II (PSII) and photosystem I (PSI) to the reaction centers (RC). The primary charge separation occurs in the RC (which works as a battery), leading to the formation of an electrochemical gradient. During the past two decades, crystallographic structures for many photosynthetic complexes (PC), including the LHCs and RCs, have been determined to a resolution of 2.5–3 Å. PCs operate in a thermal protein environment at ambient temperature and in the presence of external “classical” sources of noise. In spite of this, recent experiments based on 2D laser-pulse femtosecond photon-echo spectroscopy revealed a long-lived exciton-electron quantum coherence in PCs such as the Fenna-Matthews-Oslov (FMO) and marine algae. This occurs mainly because the dynamics of the electron transfer (ET) are so rapid (a few) that the thermal fluctuations and external noise are unable to significantly destroy quantum coherence. Consequently, the exciton/electron dynamics in LHCs-RCs must be described using quantum-mechanical methods.

An important consequence of the exciton/electron dynamics in LHCs-RCs is the high ET efficiency of the peripheral antennae complexes (close to 100%). Usually two different approaches are used to describe the influence of the protein environment on the ET. One is based on the thermal protein environment. In this case, the environment acts self-consistently on the electron system and, in combination with the transition amplitudes between sites and pigments, provides the ET rates between the sites and the Gibbs equilibrium state for the LHC-RC subsystem. The second approach is based on considering an external classical noise provided by the protein vibrations. This approach results in a transfer rate for the electron, but does not lead to Gibbs equilibrium states. The choice of approach depends on the specific experimental situation that the theoretical model is intended to describe.

We model the quantum ET in the photosynthetic RC, using a non-Hermitian Hamiltonian approach. Our model includes: (1) two protein cofactors, donor and acceptor, with discrete energy levels; and (2) a third protein pigment (sink), which has a continuous energy spectrum. Interactions are introduced between the donor and acceptor, and between the acceptor and the sink, with noise acting between the donor and acceptor. The noise is considered classically (as an external random force) and it is described by an ensemble of two-level systems (random fluctuators). Each fluctuator has two independent parameters, amplitude and switching rate. We represent the noise by a set of fluctuators with

fitting parameters (boundaries of switching rates), which allows us to build a desired spectral density of noise in a wide range of frequencies. We analyze the quantum dynamics and the efficiency of the ET as a function of: (1) the energy gap between the donor and acceptor, (2) the strength of the interaction with the continuum, and (3) noise parameters. As an example, numerical results are presented for the ET through the active pathway in a quinone-type photosystem II RC. As demonstrated in Fig. 1, the donor,  $|d\rangle$ , interacts with the acceptor,  $|a\rangle$ , through the corresponding matrix element,  $V_{da}$ . The third protein pigment, RC (sink), has a continuous energy spectrum, and is described by two parameters: (1) its density of states,  $\rho(E)$ , and (2) its strength of interaction with the acceptor,  $V_{aI}$ . The sink is described self-consistently, by using the Feshbach projection method on the “donor-acceptor” intrinsic states, within a non-Hermitian Hamiltonian approach [1]. We apply our results to the quantum dynamics of the electron transfer in the active branch of the quinone-type PSII reaction center. The collective external noise produced by the environment of the proteins acts on the donor-acceptor sub-system. Usually, the presence of noise acts as an incoherent pump in the system under consideration. But, as our results demonstrate, the simultaneous influence of both noise and the sink significantly assist, under appropriate conditions, the quantum efficiency of the electron transfer. We derived the expression for the electron transfer rate, which describes the tunneling to the sink, in the presence of noise. We calculate explicitly the corresponding region of parameters of the

noise-assisted quantum-electron transfer for sharp and flat redox potentials and for noise described by an ensemble of two-level fluctuators. Our results show that even in this simplified model, the quantum dynamics of the electron transfer to the sink can be rather complicated, and sensitive to many parameters. Further analytical research and numerical simulations are required to extend our approach for: (1) complicated dependencies of the density of states on energy in the sinks for flat and sharp redox potentials in the presence of noise and thermal environments and (2) more complicated LHCs-RCs complexes. The problem of the electron transfer optimization also requires further analysis.



*Fig. 1. A reaction center consisting of the donor and acceptor discrete energy levels, with the acceptor coupled to a sink with a continuous spectrum.*

# Exploring the Possibility of Quantitative Roles for Process Monitoring in Nuclear Safeguards

Thomas L. Burr,  
Michael S. Hamada,  
Mikhail Skurikhin,  
Brian Weaver, CCS-6

Process monitoring (PM) is increasingly important in nuclear safeguards as a supplement to material-balance (MB)-based nuclear materials accounting (NMA). The main goal for using PM is to improve the ability to detect off-normal plant operation, which could indicate intent to divert special nuclear material [1-13]. However, while PM is used to support NMA in various ways, PM does not currently provide direct input to quantitative measures of safeguards effectiveness. With this main goal in mind, programs within the DOE and the NNSA aim to advance the quantitative use of PM. In addition, analysis of the extent to which PM can provide quantitative assessment in effectiveness evaluation is one of 10 recognized technical challenges in the anticipated increased used of PM data.

**T**oward the goal of understanding how to quantify the effectiveness of PM, recent work has provided a possible new role for it, placing PM on “equal statistical footing” with NMA. The focus of our work has been on safeguards at aqueous reprocessing plants where, in the particular case of solution monitoring, PM tracks frequent measurements of bulk solution mass and volume. More recently, PM consisting of, for example, measured current and/or voltage in the electrorefiner unit in a pyro-reprocessing facility has begun to be investigated.

In this work, safeguards performance is defined as system detection probabilities (DP) under various diversion scenarios. Our proposed system includes residuals from both NMA and PM. Using somewhat limited real data to guide us, we have simulated non-anomalous and anomalous data from an aqueous reprocessing facility that includes process variation and measurement error effects. This allows the broader community full access to our perturbation methods and to generate statistically equivalent challenge data sets. Specific tasks are: (1) to provide example “benchmark” data with process variation and measurement error effects to allow the safeguards community access to the same simulated data we use to assess system DPs, (2) to provide an aqueous reprocessing facility model description in sufficient detail to enable effective expert elicitation regarding diversion scenarios, and (3) as an initial step toward model validation, to perform a “sensitivity study,” in which we estimate system DPs and assess the effect of measurement error and process variation on the estimated DPs.

As an example, Fig. 1 shows plots of simulated data from seven tanks in a generic aqueous reprocessing facility. Figure 2 shows plots of both PM and NMA residuals. The NMA residual is the usual MB. The PM

residuals arise from marking events and treating each tank as a sub-material-balance area, plus from having a model-based prediction of pulsed-column inventory and waste-stream material flows. For combined PM and NMA residuals such as shown in Figure 2, we are developing pattern recognition methods with the ability to detect special nuclear material loss (through diversion, theft, or innocent loss mechanisms such as pipe leaks) over time and/or space.

In a nutshell, we seek a more quantitative role for PM in anticipation of increased reliance on very rapid and relatively low-quality, on-line measurements. Key technical obstacles include: (1) modeling of uncertainty as related to understanding the normal background facility data with process variation and quantifying the impact of various facility misuse scenarios (to better understand the impact of different facility model fidelities on estimated safeguards performance), (2) distinguishing sensor anomalies from material loss, (3) the possibility of data falsification, and (4) developing custom pattern recognition options for combining disparate data types (from PM and NMA) on differing time scales as seen in Fig. 2.



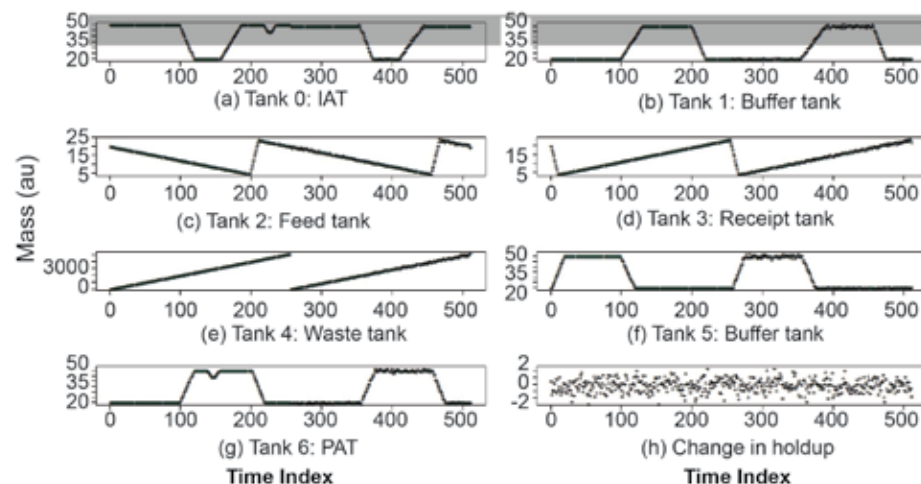


Fig. 1. Simulated data from tanks 0 to 6. The holdup in subplot (h) is in the holdup in the pulsed column between the feed and receipt tank.

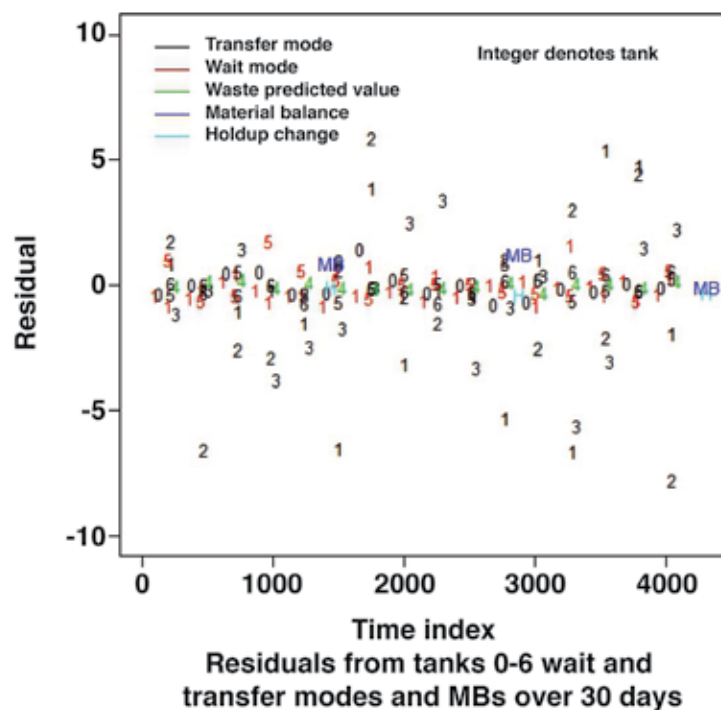


Fig. 2. Residuals from NMA and PM for the seven tanks (tanks 0 to 6) in Fig. 1.

- [1] Burr, T. et al., *J Nucl Sci Tech* **49**(2), 209 (2012).
- [2] Burr, T. et al., *Int J Res Rev Appl Sci* **8**(3), 270 (2011).
- [3] Burr, T. et al., *J of Nucl Mater Manag* **40**(2), 42 (2012).
- [4] Burr, T. et al., *J of Nucl Mater Manag*, **40**(4), 115 (2012).
- [5] Burr, T. and M.S. Hamada, *GJSFR* **11**(1), 10-30 (2012).
- [6] Burr, T. and M.S. Hamada, *Appl Radiat Isot*, **70**(8), 1675 (2012).
- [7] Burr, T. and M.S. Hamada, *GJSFR* **12**(3), 1-7 (2012).
- [8] Burr, T. et al., *Accred Qual Assur* **16**, 347 (2011).
- [9] Garcia, H. et al., *Progr Nucl Energ* **54**(1), 96-111 (2011).
- [10] Burr, T. et al., *Accred Qual Assur*, DOI 10.1007/s00769-012-0902-6 (2012).
- [11] Burr, T. and Hamada, M.S. *Nucl Instrum Meth Phys Res A*; 55-61 DOI:10.1016/j.nima.2012.05.035 (2012).
- [12] Howell, J. et al., *Comput Chem Eng* **48**, 29 (2013).
- [13] Burr, T. et al., *Stat Res Lett*, **1**(1), 6-31 (2013).

# An Innovative Parallel Cloud Storage System using OpenStack's Swift Object Store and Transformative Parallel I/O Approach

Hsing-Bung Chen,  
Benjamin McClelland, HPC-5;  
David Sherrill, HPC-3;  
Alfred Torrez, HPC-1;  
Parks Fields, HPC-5;  
Pamela Smith, HPC-3

Cloud systems (computing, service, storage, architecture, and infrastructure) are very popular in the Information Technology (IT) community today. At LANL, we have worked on high-performance computing (HPC) systems for many years. The LANL parallel log file system (PLFS) has demonstrated its superior capability for the conversion of logical N-to-1 parallel I/O operations into physical N-to-N parallel I/O operations on HPC production systems. In this article, we describe the leveraging of the scaling capability of cloud object storage systems and the transformative parallel I/O feature (Fig. 1) of the LANL PLFS and the building of a parallel cloud storage system.

**W**e used the Swift Object Store from OpenStack as our disk-based cloud storage system. OpenStack Object Storage (code-named Swift) is open-source software for creating redundant, scalable object storage using clusters of standardized servers to store petabytes of accessible data. It is not a file system or real-time data storage system, but rather a long-term storage system for a more permanent type of static data that can be retrieved, leveraged, and then updated if necessary.

We have applied a Fuse-based file system, S3QL, on top of the Switch object storage store. S3QL is an active Python/Fuse-based file system that can run on top of S3-type storage systems. These systems can transform the S3 storage system web service interface into a full-featured UNIX file system access interface. They serve as file system interfaces between the LANL PLFS and OpenStack's Switch Object Store. The proposed software architecture for this parallel cloud storage system is shown in Fig. 2.

issues. The PLFS parallel I/O scaling performance testing results are shown in Fig. 3.

As HPC systems move into the exascale computing era, the archiving exabytes of data becomes a real challenging issue. Simply buying more tapes and hard drives for storage is not a viable solution. We believe that merging advanced features from both HPC systems and Cloud systems is a promising direction.

In the future we would like to continue our research in several areas: (1) test with different data compressing algorithms; (2) examine the impact of cache size on CPU utilization, power consumption, and input/output bandwidth; (3) conduct performance testing on large-scale machines; (4) investigate other cloud storage systems (such as GlusterFS and CEPH file systems) and integrate them with the PLFS transformative parallel I/O feature; (5) enhance the S3QL with transformative parallel I/O functions; (6) apply the proposed Parallel Cloud Storage system architecture to Massive Array of Idle Disk (MAID) systems; and (7) implement dynamic power management features.

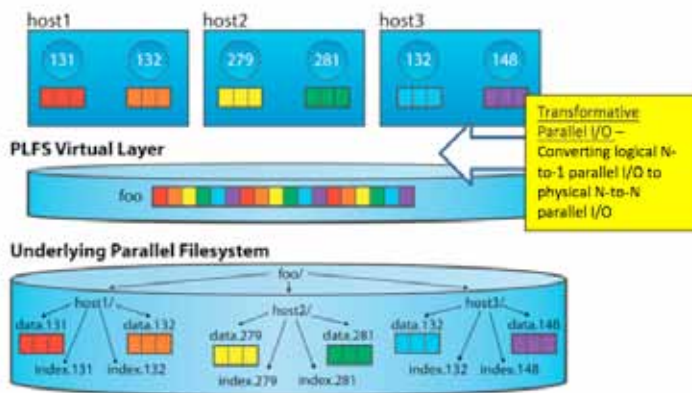


Fig. 1. Transformative parallel I/O.

We have successfully integrated the PLFS transformative parallel I/O features with the Openstack Swift Object Store. We also conducted various performance studies of I/O bandwidth and systems maintenance and management issues. We have demonstrated scaling performance results and justified performance

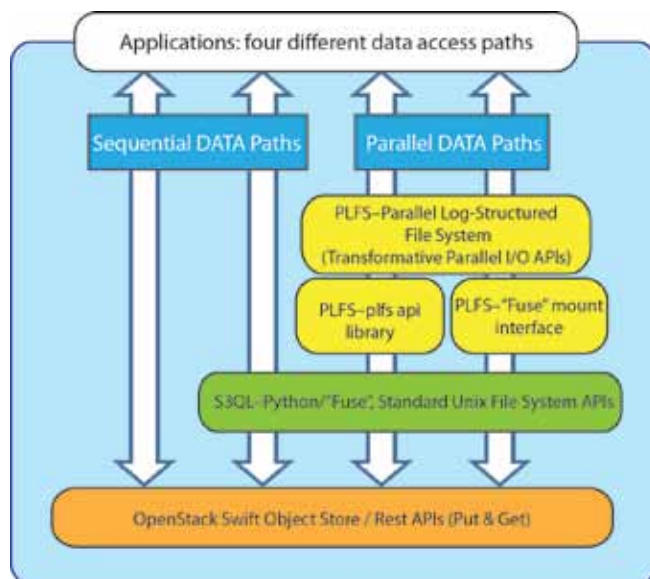


Fig. 2. Software architecture of the proposed Parallel Cloud Storage System.

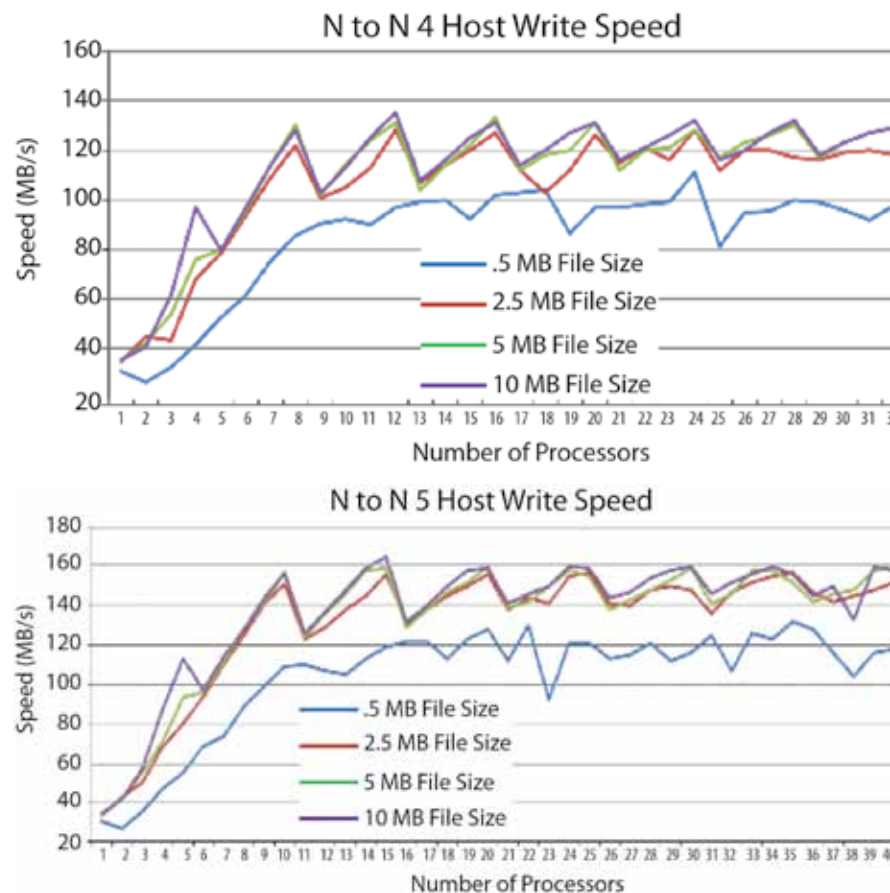


Fig. 3. PLFS parallel I/O scaling testing-multiple concurrent write session testing case.

#### Bibliography

Bent, J. et al., "PLFS: a Checkpoint Filesystem for Parallel Applications," *SC 09 Proceedings Conf High Performance Computing Networking, Storage and Analysis* (2009).

Manzanares A. et al., "The Power and Challenges of Transformative I/O," *IEEE Cluster 2012* (2012).

"OpenStack-Open Source Software for Building Private and Public Clouds," [www.openstack.org](http://www.openstack.org)

OpenStack Compute, *Starter Guide*, Openstack Document Essex Release (2012).

"S3QL-a Full-Featured File System for Online Data Storage," <http://code.google.com/p/s3ql/>, [www.rath.org](http://www.rath.org)

Armburst, M. et al., "Above the Clouds: A Berkeley View of Cloud Computing," Technical Report UCB/EECS-2009-28, University of California, Berkeley (2009).

Xiaoming, G. et al., "A Survey of Cloud Storage Systems," CSCI-B649 Cloud Computing Project Report, Indiana University (2010).

Shankaran, N. and Sharma, R. "Cloud Storage Systems-A Survey," CSCI-B649 Cloud Computing Project Report, Indiana University (2011).

Dewan, H. and R.C. Hansdah, "A Survey of Cloud Storage Facilities," *2011 IEEE World Congress on Services* (2011).



# iBridge: Improving Unaligned Parallel File Access with Solid-State Drives

Kei Davis, CCS-7;  
Xuechen Zhang, Georgia  
Institute of Technology;  
Ke Liu,  
Song Jiang,  
Wayne State University

Parallel I/O is central to high-performance computing (HPC), and improving its efficiency is an ongoing research and development problem. In a parallel I/O system, a file is striped over multiple servers or hard disks, and requests to the file are correspondingly decomposed into a number of sub-requests that are distributed over the servers. If a request is not aligned with the striping pattern, this decomposition can produce sub-requests much smaller than the striping unit. Because disks are much less efficient in serving small requests than large ones, the net throughput of the entire system can be severely degraded by serving these smaller requests or fragments. This effect can become the Achilles' heel of a parallel I/O system that is seeking scalability with large sequential accesses. The work described in this article contributed to CCS-7 research into common runtime elements for programming models for increasingly parallel scientific applications and computing platforms [1].

**W**e have implemented a scheme, iBridge, in the PVFS2 parallel file system that identifies requests for fragments and uses solid-state drives to serve them, thereby eliminating their detrimental effect on disk-based server performance. Our experimental results on the LANL Darwin cluster show that iBridge can significantly improve the I/O throughput for real-world HPC applications.

To meet the demand for high-throughput-data access on storage systems by highly parallel scientific and engineering applications, parallel file

systems such as GPFS, Lustre, and PVFS2 have been widely adopted to manage large data files such as checkpoint/restart files and the inputs and outputs of data-intensive applications. In these file systems the files are striped over a number of data servers to take advantage of aggregate network and hard disk bandwidth, while programmers are presented with a convenient linear logical file address space.

With file striping, a request for a segment of logically contiguous file space is divided into sub-requests that are distributed over

multiple data servers. While the striping unit size is usually reasonably large so that sub-requests are sufficiently large to obtain high disk efficiency at each server, the first and/or last sub-requests can be much smaller than the striping unit if the request pattern does not match the striping pattern—that is, data access is unaligned. Figure 1 depicts three such access patterns relative to the striping pattern.

We analyzed I/O traces from various computing environments and have discovered that unaligned access is common. One example is the set

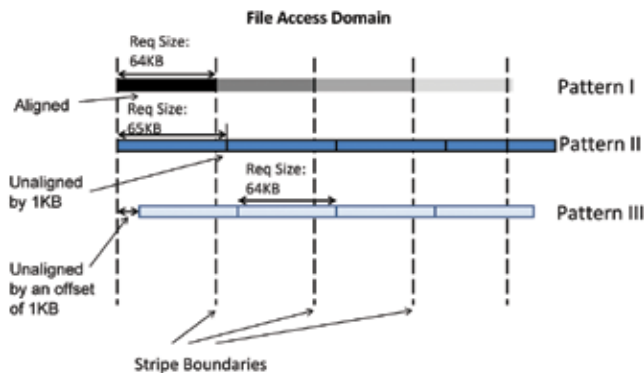
of traces of HPC applications from the Scalable I/O project at Sandia National Laboratories (SNL), including the applications ALEGRA, CTH, and S3D. As shown in Table 1, up to 62.8% (35.9% on average) of I/O requests are unaligned with the striping pattern on the data servers if we assume a 64 KB striping unit, the default for PVFS2.

**Table 1.** Percentages of unaligned and random data accesses in different I/O traces with a 64 KB striping unit. Unaligned refers to requests that are larger than a striping unit (64 KB) but are not aligned to the striping unit boundaries.

App	Unaligned (%)	Random (%)	Total (%)
ALEGRA-2744	35.2	7.3	42.5
ALEGRA-5832	35.7	6.9	42.6
CTH	24.3	30.1	54.4
S3D	62.8	5.8	68.6

These smaller sub-requests, which we call fragments, are effectively random accesses on their respective servers. Because a hard disk can be less efficient in serving random requests than sequential ones by one or more orders of magnitude, fragments and other sub-requests of the same request can be served with very different efficiency. In the case of synchronous requests the entire storage system's productivity is degraded.

To investigate the effects of unaligned access on storage system performance we ran the *mpi-io*-test benchmark in which N processes iteratively read data from a 10-GB file striped over eight data servers. Figure 2 shows that throughput for aligned access is approximately twice that for unaligned access.



**Fig. 1.** File access with three different alignment patterns. In Pattern I, access is perfectly aligned. In Pattern II, sub-requests are greater than the striping pattern size. In Pattern III, the sub-requests are the same size as the striping unit but are offset, a common situation for file formats that contain header data.

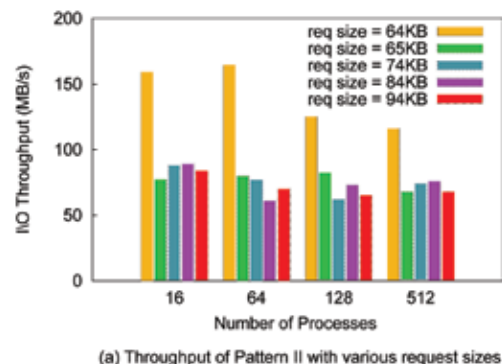


Fig. 2. I/O throughput for *mpi-io*-test for access pattern II from Fig. 1 for various request sizes. Note that for a request size of 64 KB access is perfectly aligned.

- What criteria distinguish regular random requests and fragments?
- What are the criteria for deciding whether a regular random request or a fragment should be admitted into the SSD?
- Because SSD space is limited, how should it be allocated between regular random requests and fragments?

iBridge was prototyped in the PVFS2 2.8.2 parallel file system on the Darwin cluster at LANL. The cluster includes 116 48-core 2-GHz AMD Opteron nodes interconnected with a dual-rail 4X QDR Infiniband network. We configured eight nodes as data servers and one node as the meta-data server. Each data server had one 7200-RPM disk drive and a 120-GB SSD.

We sought experimental answers to the following questions, among others.

- Can unaligned requests be effectively served using iBridge in hybrid storage systems?
- Is iBridge effective for real scientific workloads with diverse data access patterns?
- How much overhead does iBridge add to the PVFS2?

iBridge was evaluated using both a range of synthetic benchmarks such as *mpi-io*-test and real-world scientific

The objective of iBridge is to use solid-state drives (SSD) to serve fragments produced by unaligned data accesses. Such a use of SSDs represents a highly desirable combination—the SSD's strong performance advantage for random access, the fragments' relatively small sizes, and a strong need to efficiently serve fragments. However, to make the approach truly effective we had to solve a number of algorithmic problems:

applications such as those given in Table 1, in the latter case by replaying their I/O traces rather than actual execution of the applications. Figure 3 shows the throughputs for *mpi-io*-test for both reads and writes, with and without iBridge, for 65-KB requests as a function of number of processes. In these cases, iBridge provides more than a factor of two performance improvement. Table 2 shows the service times, with and without iBridge, when replaying the HPC applications from the SNL Scalable I/O project for which alignment data was given in Table 1. With iBridge, the request service times are reduced by 13.9%, 18.7%, 25.9%, and 29.8%, respectively. Other measurements showed that iBridge added negligible time and space overhead to the PVFS2 system.

Parallel I/O is a cornerstone of high-performance scientific computing and continued advances in its performance are needed to keep up with advances in the rest of the computing platform. With iBridge we have shown that with a very modest investment in relatively new technology—the solid-state drive—and appropriate algorithmic development, we can significantly improve the performance of existing parallel I/O systems—and with negligible additional overhead and no changes to the application developer's interface to it.

**Table 2.** Comparison of request service times when for I/O traces replayed with and without using iBridge.

	ALEGRA. 2744	ALEGRA. 5832	CTH	S3D
PVFS2	16.6ms	17.2ms	19.4ms	36.0ms
PVFS2 + iBridge	14.2ms	14.0ms	14.4ms	25.3ms

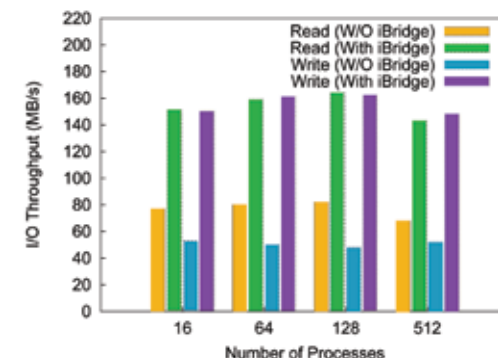


Fig. 3. Throughputs for *mpi-io*-test with and without iBridge as process count increases.

[1] Zhang, X. et al., "iBridge: Improving Unaligned Parallel File Access with Solid-State Drives," LA-UR-12-25294; *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, (2013).

# Improving Disk Performance: A Prefetching Scheme Exploiting Data Layout and Access History

Kei Davis, CCS-7;  
Xiaoning Ding,  
Xiaodong Zhang, The Ohio  
State University  
Song Jiang, Wayne State  
University

Prefetching is an important technique for improving effective hard disk performance. It attempts to accurately predict the data to be requested and load it ahead of the arrival of application requests. Current disk prefetch policies in major operating systems (OS) track access patterns at the level of file abstraction, which has limitations and so cannot realize the full performance improvements achievable by prefetching. We have designed a system, DiskSeen, that performs prefetching directly at the level of disk layout, and in a portable way. An important design consideration is that our technique is entirely supplementary to, and works synergistically with, any present file-level prefetch policies. Our implementation in the Linux kernel shows that it can significantly improve the effectiveness of prefetching, reducing execution times by 20–53% for micro-benchmarks and real applications. Even with workloads specifically designed to expose its weaknesses DiskSeen exhibits only minor performance loss.

Current disk prefetch policies in major OSs track access patterns at the level of file abstraction. While this is useful for exploiting application-level access patterns, there are two reasons why file-level prefetching cannot realize the full performance improvements achievable by prefetching: (1) certain prefetch opportunities can only be detected by knowing the data layout on disk, such as the contiguous layout of file meta-data or data from multiple files; and (2) non-sequential access of disk data (requiring disk head movement) is much slower than sequential access, and the penalty for mis-prefetching a “random” block (unit of disk capacity), relative to that of a sequential block, is correspondingly more costly. To overcome the inherent limitations of prefetching at the logical file level we propose to perform prefetching directly at the level of disk layout and in a portable way. Our technique, called DiskSeen, is intended to be supplementary to, and to work synergistically with, any present file-level prefetch policies.

In essence DiskSeen is a sequence-based, history-aware prefetch scheme based on two observations: (1) accesses of disk blocks in a particular order are likely to be repeated; and (2) because it is performed asynchronously with application execution, if prefetching is not so inaccurate as to interfere with application progress, it is essentially free.

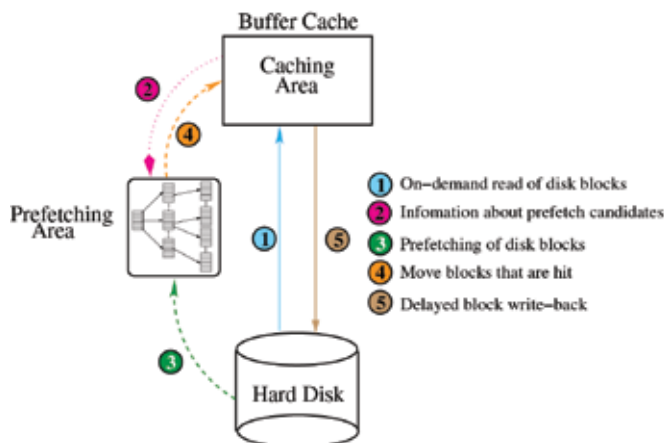


Fig. 1. A simplified diagram of the DiskSeen system.

DiskSeen tracks the locations and access times of disk blocks, and based on analysis of their temporal and spatial relationships, seeks to improve the sequentiality of disk accesses and overall prefetching performance. It also implements a mechanism to minimize mis-prefetching to mitigate the corresponding performance penalty.

We leave file-level prefetching enabled; DiskSeen concurrently performs prefetching at a lower level to mitigate the inadequacies of file-level prefetching. DiskSeen seeks to detect sequences of block accesses based on block disk addresses, or logical block numbers (LBN). At the same time, it maintains block access history and uses the history information to further improve the effectiveness of prefetching when recorded access patterns are observed to be repeated.

There are four objectives in the design of DiskSeen: (1) Efficiency—we ensure that prefetched blocks are in a localized disk area and are accessed in ascending order of their LBNs for optimal disk performance; (2) Eagerness—prefetching is initiated immediately when a prefetching opportunity emerges; (3) Accuracy—only the blocks that are highly likely to be requested are prefetched—significant inaccurate prefetching automatically suppresses prefetching; and (4) Aggressiveness—prefetching is made more aggressive if it helps with accuracy and reduces request service times. Conversely, if it is detected to be increasing service times (because of inaccurate predictions) it is throttled back.

Buffer cache is divided into prefetching and caching areas according to their roles in the scheme (see Fig. 1). A block could be prefetched into the prefetching area based on either current or historical access information—both are recorded in the disk block table, or as directed by



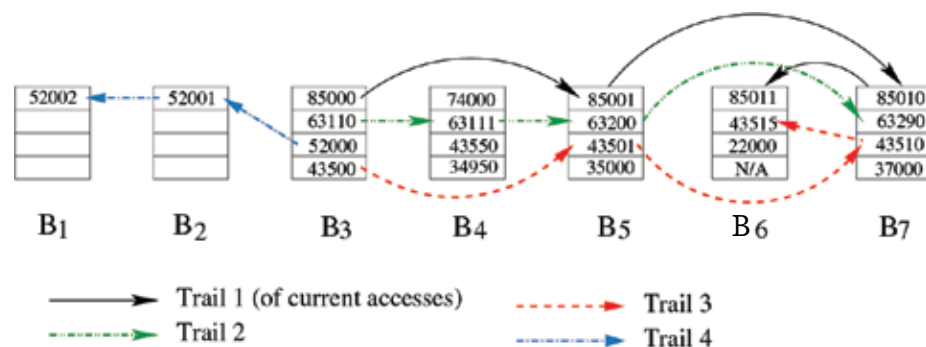


Fig. 2. Access trails. B1 through B7 are consecutive contiguous blocks in the block table. There are four trails starting from block B3: one current trail and three history trails. Trail 1 (B3, B5, B7, B6) corresponds to the ongoing continuous block accesses.

file-level prefetching. The caching area corresponds to the traditional buffer cache and is managed by the existing OS kernel policies except that prefetched but not-yet-requested blocks are no longer stored in the cache. A block is read into the caching area either from the prefetching area, if it is hit there, or directly from disk, all in an on-demand fashion.

A central component of DiskSeen is its mechanism for maintaining access sequence history. To describe access history we introduce the term trail to describe a sequence of blocks that has been accessed with a small time interval between each consecutive pair of blocks in the sequence and is located in a bounded region. DiskSeen maintains a history of previously seen access sequences—trails—to provide its predictive capability. An important point is that a trail may be any sequence of blocks, whereas conventional disk-level prefetching typically relies the detection of strictly sequential access of contiguous blocks. Figure 2 depicts a segment of the trail history data structure.

Figure 3 shows the execution times for a selection of benchmarks (described in [1]) chosen both to showcase DiskSeen and to specifically thwart DiskSeen's predictive capability. Two observations are that (1) even when DiskSeen is wholly ineffective, it does not significantly hurt performance; and (2) an application need not necessarily be run more than once for the history mechanism to be useful.

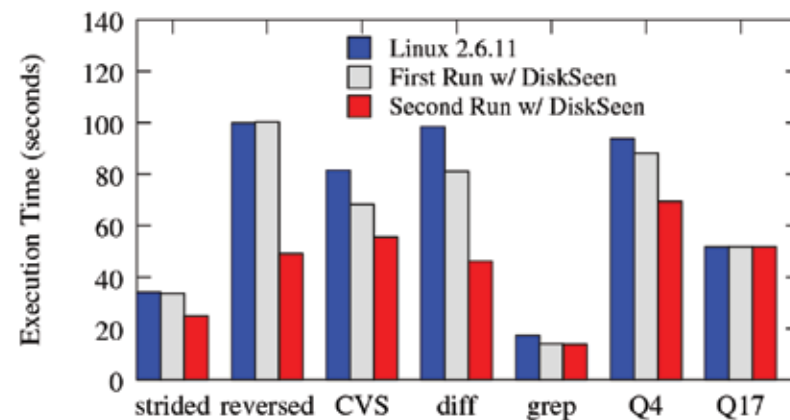


Fig. 3. Execution times of six benchmarks. Linux 2.6.11 refers to the stock Linux kernel.

[1] Jiang, S., et al., "A Prefetching Scheme Exploiting both Data Layout and Access History on Disk," *ACM Trans Storage*, to appear (2013).

# Making Memory Swapping Practicable: Synergistic Coupling of SSD and Hard Disk for QoS-Aware Virtual Memory

Kei Davis, CCS-7;  
Song Jiang, Wayne State  
University;  
Xuechen Zhang, Georgia  
Institute of Technology;  
Ke Liu, Wayne State  
University

Hard disk space has long been used to provide a virtual extension of main memory in order to allow computer programs with memory requirements that are greater than the available main memory to run. For large parallel machines, however, this mechanism is often not made available because of the excessive performance penalty—disk access is orders of magnitude slower than main memory. The relatively recent advent of the flash-memory-based solid-state drive (SSD), with access time several times faster than disk, would seem to be a technological improvement for this purpose. However, because SSD has sharply limited write endurance, a virtual memory extension based on SSD could quickly become unreliable. We have developed a quality of service (QoS)-aware system [1] that uses SSD and disk in tandem, exploiting the relative strengths of each and which provides a virtual memory extension that can be as fast or faster than SSD alone while minimizing the number of writes to SSD.

When a computer program's memory requirement, which typically grows during program execution, exceeds available memory, there are two possible outcomes depending on how the system is configured. The operating system on typical desktop computers or small servers transfers the least-recently-used pages of the memory image to hard disk to free up memory space, and transfers them back into memory (at the expense of moving other pages out) if they are later needed. This process is known as swapping, and the effective increase of available memory is known as a virtual memory extension. This is a cost-effective solution in the sense that the cost and energy consumption of disk, per unit of capacity, is orders of magnitude less than main memory (DRAM).

The price for this nearly free extra memory is performance—disk access, both in terms of latency (waiting time until transfer starts) and bandwidth (rate of data transfer once underway) is orders of magnitude slower than DRAM. In some scenarios this is acceptable—better that the program run slowly than quit running altogether. For large-scale scientific computing, however, this may not be acceptable—the computing platforms are expensive to purchase and to operate and maximizing their throughput is an economic imperative. As such, a swapping mechanism in the operating system is typically not available.

The relatively recent advent of the flash-memory-based solid-state drive (SSD), with much greater performance than disk for random and small accesses, suggests that by using SSD rather than disk the use of virtual memory extension would be acceptable and useful in more scenarios than it is currently. There are two problems with this simplistic approach. The first is that while SSD is much less expensive than DRAM in terms of capacity, it is more expensive than disk. The second is that SSD

has a strictly limited life expectancy in terms of the number of write operations it can endure and, to make matters worse, the technological trend is towards greater capacity and lower cost at the expense of write endurance.

Our idea was to consider the relative strengths and weaknesses of disk and SSD, and to develop an algorithm that would distribute the load over an SSD-disk pair that would exploit their relative strengths to provide a virtual memory extension that is highly performant, cost effective, and with life expectancy comparable to disk.

**Table 1.** Relative characteristics of DRAM, SSD, and disk.

	DRAM	SSD	DISK
Power/GB	high	low	med
Cost/GB	high	med	low
Random/short access performance	high	med	low
Sequential access performance	high	med	med
Write endurance	unlimited	limited	unlimited

Table 1 compares the relevant characteristics of DRAM, SSD, and disk. With respect to power with other considerations aside, disk is a more efficient storage medium than DRAM, and SSD even more so. Regarding cost per unit of capacity, disk is less expensive than SSD, suggesting that the greater part of the memory extension would best reside on disk. SSD is much faster than disk for random (non-sequential) and

short accesses—this is because a non-sequential access will typically require a disk head movement and then wait for the data location on the spinning disk to appear under the disk head, requiring several milliseconds, before data starts to be read. The corresponding latency for SSD is tens of microseconds.

Where disk performance is more comparable to SSD is for long sequential accesses. Here the disk head can read or write data continuously as the disk spins underneath it, requiring only a short movement to the next concentric track after spanning the current one. In this scenario using disk instead of SSD is preferable for two reasons—cost/capacity (long reads or writes) and longevity (for writes).

The first step was to investigate whether representative memory-intensive applications generate memory-page access patterns that would warrant the use of disk to significantly reduce SSD writes without significant performance loss compared to SSD-only swapping. An important observation is that a sequence of page accesses need not be strictly sequential to be cast as a sequential access to disk. For example, if a sequence of accesses covers (possibly with small gaps) a contiguous span of memory pages it may directly cast as sequential if the pattern is detected the first time and is repeated. Extending this further, if any sequence of memory accesses is repeated it may be translated to a sequential access to disk. Instrumentation of representative applications shows that such patterns are common.

Based on these ideas and observations we developed a system in the Linux kernel, HybridSwap, and conducted an extensive performance evaluation using representative benchmarks. The behavior of HybridSwap is dynamically tunable—greater or lesser performance can be obtained by biasing traffic toward or away from the SSD, but with correspondingly greater or lesser write access to the SSD. There are two immediate consequences, the more obvious being that wear on the SSD can be controlled. The second is that we can provide quality-of-service functionality whereby

bounds on the swapping penalty (within the capacity of the swapping system) may be specified for individual applications by effectively prioritizing their use of SSD.

Here we show a sample of results with HybridSwap at its default settings, chosen to yield application performance on par or slightly better than using SSD alone. Table 2 shows that HybridSwap reduces SSD writes by 16% to 40% for disparate benchmarks, and with run time reductions ranging from -0.8% to 7.1%.

**Table 2.** Reduction in writes to SSD when running multiple competing instances of the Memcached (Memc), ImageMagick (Image), Matrix Inverse (Matrix), and Correlation Computation (CC) benchmarks.

	Memc	Image	CC	Matrix
Write reduction	37%	40%	22%	16%
Performance improvement	-0.8%	4.6%	7.1%	0.5%

[1] Zhang, X. et al., “Synergistic Coupling of SSD and Hard Disk for QoS-aware Virtual Memory,” *IEEE International Symposium on Performance Analysis Systems and Software (ISPASS)* (2013).



# Reliability Models for Double Chipkill Detect/Correct Memory Systems

Nathan DeBardeleben,  
Sean Blanchard, HPC-5;  
Rakesh Kumar,  
Stevenson Jian,  
University of Illinois;  
Vilas Sridharan, Advanced  
Micro Devices

Chipkill correct memories are an advanced type of error correction memories used on many modern high-performance computing (HPC) systems and high-end servers. While previous work has shown chipkill to be extremely beneficial, the existing analytical reliability models have focused on detecting and correcting a single bad symbol per codeword. However, chipkill technology that can detect and correct two bad symbols per codeword already exists. In this work, we propose a reliability model for double chipkill detect/correct memory (DCC) systems. Additionally, we present a Monte Carlo simulation that tracks very well with the analytical model. We use this to look at what this model says about meantime to error as an HPC system ages with several different memory types. We also examine how differing chipkill technologies impact the meantime to replacement of a dual in-line memory module (DIMM).

Chipkill correct is an advanced type of error correction in memory that significantly improves the reliability of memory by allowing continued memory operation in the event of device-level failures in memory. Large-scale studies show that chipkill correct reduces the uncorrectable error rate of memory by  $10\times$  [1] to  $42\times$  [2] compared to Single Error Correction, Double Error Detection (SECDEC) error correcting codes (ECC). As the size of memory continues to increase, the demand for higher reliability in memory increases as well. As a result, chipkill correct memories have become very popular among HPC systems and high-end servers with large memory capacity.

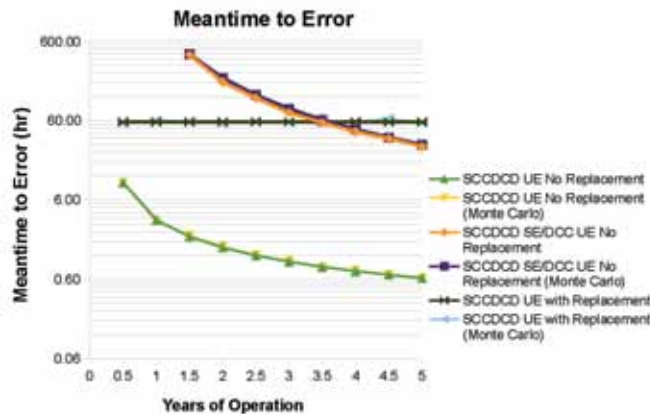


Fig. 1. Comparison of the mean time to different types of errors as predicted using the analytical models to that observed by Monte Carlo simulation for the evaluated memory organization.

As chipkill correct memories become commonplace, there will be a greater desire from system designers to predict the reliability of their system for a particular strength of chipkill correct memories, a particular memory organization such as capacity per DIMM and ranks per channel, etc., and a desired maintenance schedule of how often to perform memory scrubbing and to replace failing DIMMs in memory. Existing measurement of uncorrectable error rates for chipkill correct memories [1,2] are for specific memory organizations and therefore do not provide a way to extrapolate the reliability of memory for different memory organizations. Although a large body of work on the correctable error rate of error protected memory exist in literature [3-5], they only consider uncorrectable error rates on an individual codeword by codeword basis and ignore the correlation of faults between codewords due to the device-level faults such as row, column, and subbank faults. However, the driving factor for the increased reliability of chipkill correct

memories is that they can correct against device-level faults that affect a large number of cells in memory per fault, some of which are not correctable by SECDED. As a result, by ignoring device-level faults, previous work on the reliability of ECC memories is adequate for chipkill correct memories.

The purpose of our work is to model the reliability of chipkill correct memories for an arbitrary memory organization and to calculate how often DIMMs need to be replaced for a certain type of chipkill correct memory. Such information will help guide the design of the next generation of HPC systems.

Our work creates an analytical model of memory. We make many simplifying assumptions and are careful to explain why these assumptions are realistic and still make the results useful and interesting. The model allows us to calculate the probability of uncorrectable error SCCDCD, probability of silent data corruption for SCCDCD, probability of uncorrectable and silent data corruption for DCC, and the probability of error with DIMM replacement. These involve many complex equations that are not appropriate for this publication and are not given here.

In addition to our analytical model, we have also designed a Monte Carlo simulation. The Monte Carlo simulation considers a single channel at a time. Each run of Monte Carlo simulates how long it takes for the corresponding condition (replacement for SCCDCD or DCC or uncorrectable fault for SCCDCD) to occur. In each simulated time interval, a random floating point number is generated for each type of fault. If a random number is less than the probability of its corresponding type of fault for the current time interval, a fault of that type is injected into the simulated memory channel. When the condition of unrecoverable error (UE)/SE/replacement is met, the total time span between the start of the simulation to the failing time interval is reported. After a number of time intervals, whose total span equals a scrubbing period,

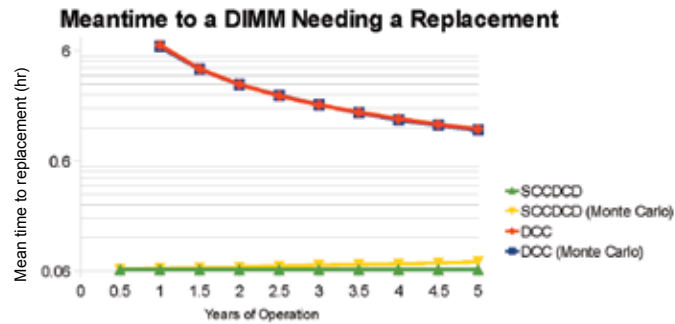


Fig. 2. Comparison of the mean time to a DIMM needing replacement as predicted using the analytical models to that observed by Monte Carlo simulation for the evaluated memory organization.

have passed, all the row and symbol faults are cleared to simulate the effect of scrubbing. Similarly, after a number of time intervals, whose total span equals a DIMM replacement interval, have passed, all the remaining faults are cleared to simulate the effect of DIMM replacement. The total number of Monte Carlo runs is equal to the total number of channels in the target memory system. The outputs from each run are grouped into different bins depending on which time interval they fall into.

**Table 1. Simulated Memory Organization**

Size of Row in a Bank	8 kilobytes
Banks Per Rank	8
Capacity Per Device	2 gigabytes
Data Devices Per DIMM	32
DIMMs Per Channel	2
Total Memory Size	64 petabytes
Number of Scrubs Per Week	1

Table 1 shows the memory organization used for evaluation. Note that we are simulating an early exascale system design, so newer simulations would exacerbate DIMM errors due to larger capacities of more recent exascale system design predictions. Figure 1 compares the mean time to different types of errors as predicted using the analytical models to that observed by Monte Carlo simulation for the evaluated memory organization. A couple of important observations can be made from Fig.1. First, the comparison between SCCDCD UE with no replacement versus SCCDCD SE/DCC UE with no replacement show that the reliability of tolerating two bad symbols is almost two orders of magnitude greater than the reliability of tolerating only one

bad symbol per codeword. The second observation derived by comparing SCCDCD UE with replacement to SCCDCD UE without replacement is that with DIMM replacement the meantime to uncorrectable error can be dramatically increased even with infrequent DIMM replacement (the replacement schedule is once per week as given in Table 1). In other words, DIMM replacement need not be immediate and but can be done at a convenient pace. Finally, DCC may not require any DIMM replacement until much later in the operation of the machine.

Figure 2 compares the mean time to a DIMM needing replacement as predicted using the analytical models to that observed by Monte Carlo simulation for the evaluated memory organization. The visible difference between the Monte Carlo model and the analytical model for SCCDCD is due to the fact that in the Monte Carlo model a DIMM that needs to be replaced is taken out of the total pool of DIMMs instead of actually being replaced, which causes the total number of DIMMs simulated to decrease over time. Since a relatively large number of DIMM replacements are required, the deviation from the analytical model can become large. On the other hand, since the number of DIMM replacements is very small for DCC, the deviation from the analytical model is not noticeable.

We are taking this work in several directions. On current systems the community does not have a good understanding of DIMM device faults and how that should drive DIMM replacement strategies. This can result in either a failure prone environment or be wasteful in disposal of (as we have shown here, under some circumstances) usable DIMMs. Furthermore, we are working to model actual replacement strategies including the amount of time it takes to replace a DIMM and the cost associated with replacement so that we can make some very practical recommendations. Additionally, we are looking at some next generation triple chipkill technologies and developing representative models and simulations. We are working with AMD and SNL to compare results from DRAM errors on Cielo with the results shown in [2].

- [1] Schroeder, B. et al., "DRAM Errors in the Wild: A Large-Scale Field Study," *Sigmetrics* ACM, New York, NY 193 (2009).
- [2] Sridharan, V. and D. Liberty, International Conference on High Performance Computing Networking, Storage and Analysis, Article 76, Supercomputing 2012.
- [3] Mkherjee, S. et al., "Cache Scrubbing in Microprocessors: Myth or Necessity?" *Pacific Rim International Symposium Dependable Computing (PRDC)* (2012).
- [4] Saleh, A.M. et al., *IEEE Trans Reliab* **39**(1), 114 (1990).
- [5] Schiano, L. et al., "Markov models of fault-tolerant memory systems under SEU," *IEEE Workshop Memory Technology, Design, and Testing* (2004).

# Network Traffic Generator for Cyber Security Testbeds

Hristo Djidjev, CCS-3;  
Lyudmil Aleksandrov,  
Bulgarian Academy of  
Sciences

We have developed an algorithm for generating secure shell (SSH) network traffic that can be used as a test bed for evaluating anomaly detection and intrusion detection tools in a cybersecurity context. Given an initial dataset describing real network traffic, the generator produces synthetic traffic with characteristics close to the original. The objective is to capture complex relationships between hosts (who communicates with whom) and between sessions—such as which sessions are interrelated in time and weekly and diurnal usage patterns.

**D**etecting malicious software and intrusions in computer networks has turned into a task of utmost importance for cyber professionals and is a major challenge for researchers [1]. One popular approach for detecting intrusions is to analyze the network traffic and find patterns of anomalous behavior [2]. Cybersecurity tools based on anomaly detection typically involve two phases. In the first, the training phase, historical data is statistically analyzed and patterns of normal behavior are extracted from the data. Specifically, such patterns may include the statistical distributions of some parameters of the traffic, for example, its volume or the session durations, and thresholds of the maximum deviation from the averages beyond which the traffic will be considered abnormal. In the second, detection phase, current traffic is analyzed and compared against the model of normal traffic and any session that deviates from that model is labeled as anomalous and is further investigated by domain experts.

In order to test existing and new anomaly and intrusion detection systems it is often preferable to use a generator that can produce network systems, traffic that resembles real observed traffic [3] and that uses adjustable parameters such as the number of hosts, number of sessions, and average session durations. Using real traffic for testing purposes has several drawbacks—network data can be difficult or impossible to get for many researchers, and sharing data between researchers can pose security or privacy issues or be forbidden by cyber policies. For anomaly detection purposes, it will be hard or impossible to distinguish between normal and malicious traffic and therefore hard to evaluate the performance of the tested anomaly detection tool. Using a traffic generator allows one to mix normal and malicious traffic in various scenarios and also vary the other traffic parameters so that the anomaly detector can be tested under different conditions.

In this report we describe our work on generating normal (as opposed to malicious) network traffic based on the SSH protocol. The traffic

data we generate is in the form of a set of SSH sessions, where for each session we provide its source and target nodes (corresponding to the IP addresses of hosts in the real traffic), its start time, and its duration. Our goal is to capture patterns of real traffic associated with relationships between pairs of hosts and pairs of sessions. Specifically, we are interested in interrelated sessions that may have been started by the same user during the same activity. For instance, a legitimate user may start at host A, login to host B, go from host B to host C (as host C may not be directly reachable from A), and then close all sessions. In such a case we consider sessions (A,B) and (B,C) related. We merge such interrelated sessions in structures we call “telescoping subgraphs” (TSG) [4], see Fig. 1. In the example, the TSG is a graph consisting of three nodes, A, B, and C, and two edges (A,B) and (B,C). Intruders, unlike legitimate users, may have gained access to node A and from there can try to explore the network by hopping from host to host, forming TSGs with entirely different characteristics [5].

The difference between the patterns of legitimate users and intruders has proved useful and has been exploited in previous anomaly detectors [4,5]. Hence, one of our objectives is to generate traffic whose hosts and sessions are interrelated in a way that resembles the one observed in the original traffic. Other objectives are to preserve the temporal aspects of the traffic—namely, the different levels of activity during different days of the week and different times during the day and the session durations, the total volume of the traffic, and the distributions of the node outdegrees and indegrees, defined as the numbers of sessions with origin or that target a given host, respectively.

Our algorithm consists of two phases. During the analysis phase, the real network traffic is analyzed and a profile is computed that summarizes traffic properties that are essential for the generation process. The specific parameters of the original traffic that we compute and store for G are stored in the following data sets: (D1) the distribution of the sizes

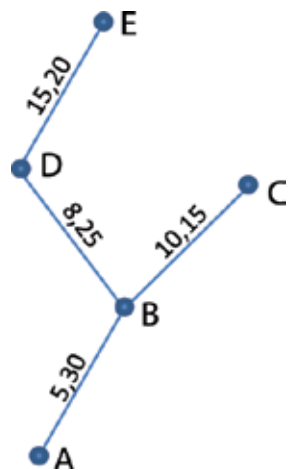


Fig. 1. A TSG. Edges have labels (s,e); S is the start time and t is the end time of the corresponding sessions. Two edges with labels (s1, e1) and (s2, e2) can appear consecutively in a TSG only if they satisfy  $s1 \leq s2$  and  $e1 \geq e2$  (the telescoping property).



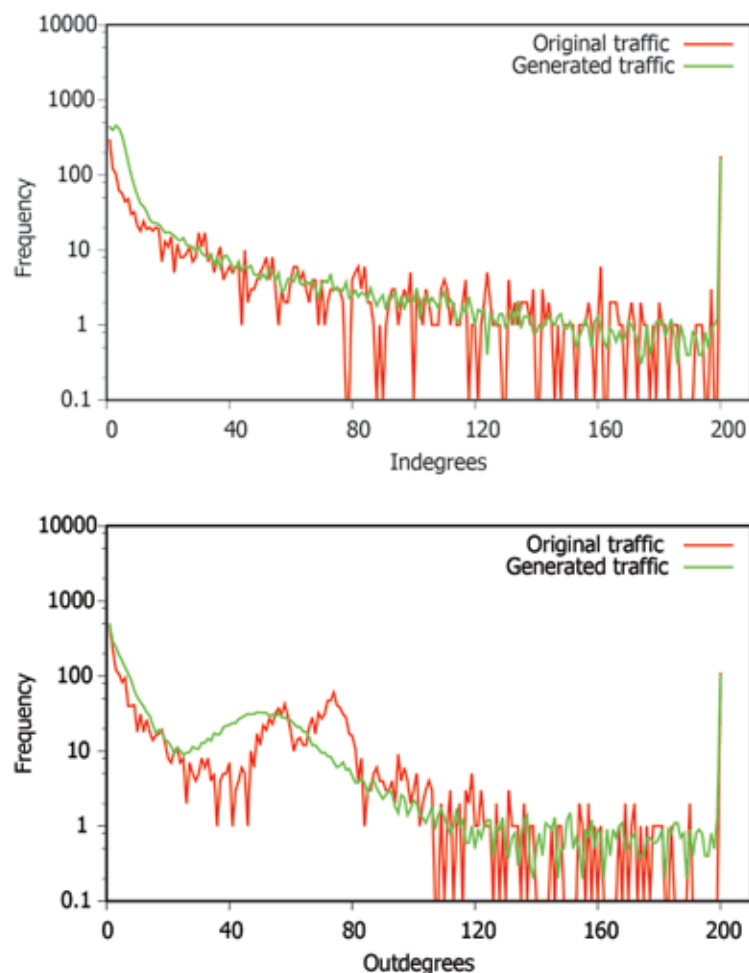


Fig. 2. Plots showing the indegree and outdegree frequencies for the original and the generated traffic. In both cases, the traffic is over a period of 28 days, but the generated-traffic data are averaged over 10 randomly generated instances.

and the total number of TSGs, (D2) the frequency of the edges (how many times each edge has been observed), (D3) the frequency of length-2 subpaths of all TSGs, and (D4) the average duration for each edge. During the generation phase, given a time window  $W$ , the profile generated in the first phase is used to generate new traffic in  $W$ . For each weekday  $w$  and each hour  $h$ , the data in the profile associated with the pair  $(w,h)$  is used. This allows us to more accurately capture the diurnal patterns of the traffic.

In order to test the method, we computed several characteristics of the original and the generated traffics and compared them. The original data were extractions of LANL network traffic collected over a period of several years. We compared TSG sizes, indegrees and outdegrees of the nodes (the hosts), and the session durations. Since our generation algorithm uses the TSG sizes information during the generation process, it was natural to expect that the TSG sizes in the original and the generated traffic would be very close. This was confirmed by the experiments. But information about the indegrees and the outdegrees was not used explicitly during the generation, so it was interesting to see if there is any correlation between these characteristics. Figure 2 shows the results of that comparison. There is a very good correlation as well as some small differences (perfect correlation is not desirable as it would indicate over-fitting). The data for the plots for the new traffic were averaged over 10 runs, which is one of the reasons they are smoother and have fractional values (between 0 and 1). The analysis of the session duration data shows a similar good correlation.

- [1] Dunlavy, D.M. et al., SNL report SAND 2009-0805 (2009).
- [2] Garcia-Teodoro, P. et al., *Comput Secur* **28**, 18 (2009).
- [3] Sommers, J. et al, Computer Science Department Technical Report 1525, University of Wisconsin (2006).
- [4] Djidjev, H. et al., "Graph Based Statistical Analysis of Network Traffic," Ninth Workshop on Mining and Learning with Graphs, San Diego, CA, August 20-21, (2011).
- [5] Neil, J.C., "Scan statistics for the online detection of locally anomalous subgraphs," Ph.D. Thesis, University of New Mexico (2011).

# PENNANT: A Research Tool for Unstructured Mesh Physics on Advanced Architectures

Charles R. Ferenbaugh,  
HPC-1

Novel computer architectures such as graphics processing units (GPU), many-core chips, and IBM BlueGene are becoming common in the high-performance computing (HPC) world. Existing physics application codes require major modifications to perform well on these architectures. LANL has developed the PENNANT mini-app as a research tool for finding efficient implementations of unstructured mesh physics algorithms. It contains mesh data structures and a few physics algorithms adapted from the LANL shock physics code FLAG and will provide insights on how to optimize FLAG and other similar codes for future architectures.

The HPC world is entering a major transition. Novel architectures such as GPUs, many-core chips, and IBM's BlueGene are becoming common in supercomputer clusters. Architectures such as these provide high computational performance combined with low power usage, and are likely to be used in future systems such as the "exascale" systems being discussed in the international HPC community.

These systems will pose significant challenges to all scientific software developers—introducing new programming models to manage the increased hardware complexity and requiring major rewrites of existing software. They will be particularly challenging for developers of algorithms for general unstructured meshes—that is, computational meshes containing arbitrary polygons (in 2D) or polyhedra (in 3D). An example is shown in Fig. 1. These meshes have irregular connectivity and memory usage patterns, making them more difficult to work with

than meshes with more regular structure. As a result, unstructured mesh methods tend to lag behind other types of physics methods in advanced architecture research.

However, since unstructured mesh codes are common at LANL and elsewhere, it is important to find ways to run such codes efficiently on these architectures. As a tool for research in this area, LANL has developed a small application, or mini-app, called PENNANT [5].

PENNANT contains approximately 2200 lines of C++ source code. It implements a small subset of the physics of the LANL shock physics application FLAG. Like FLAG, it operates on general unstructured meshes (meshes

containing arbitrary polygons). It currently has implementations for serial and multi-core processors—the multi-core version uses the standard OpenMP programming model. A GPU implementation using the CUDA programming language is in progress, building on previous research [4].

PENNANT provides the following basic physics capabilities from FLAG:

- Lagrangian staggered grid hydrodynamics (SGH) [2] for basic fluid flow
- Single material, gamma-law gas equation of state
- Temporary Triangular Subzoning (TTS) [3] for subzonal pressures
- Campbell-Shashkov tensor artificial viscosity [1]

These capabilities are sufficient to run a few simple test problems in shock physics.

The following PENNANT timing results were obtained on the Darwin research cluster at LANL. The nodes of this cluster have 4 12-core AMD Opteron 6168 CPUs, for a total of 48 cores, each running at 1.90 GHz. The nodes used in this study also have NVidia M2090 GPUs attached.

Two different versions of the Noh test problem [6], nohsquare and nohpoly, were used. A sample output from nohpoly is shown in Fig. 2.

Figure 3 shows timings from a scaling study done using the OpenMP version with varying numbers of threads. The dashed line shows ideal scaling of the serial version, for comparison. For both problems, the OpenMP implementation of PENNANT scales well on up to 32 cores, but starts to level off when using the full 48 cores of a Darwin node. The 32-core runs showed a speedup of 22× over the serial version.

A GPU implementation is in progress. Currently the force calculation is running on the GPU, and timings on this portion of the code show about a 14× to 17× speedup (problem-dependent) over the corresponding serial code. Past experience suggests that the full GPU version, when complete, will show a similar speedup.

PENNANT demonstrates that unstructured mesh physics can be implemented efficiently on multicore processors and GPUs. It also shows



Fig. 1. Example of an unstructured mesh.

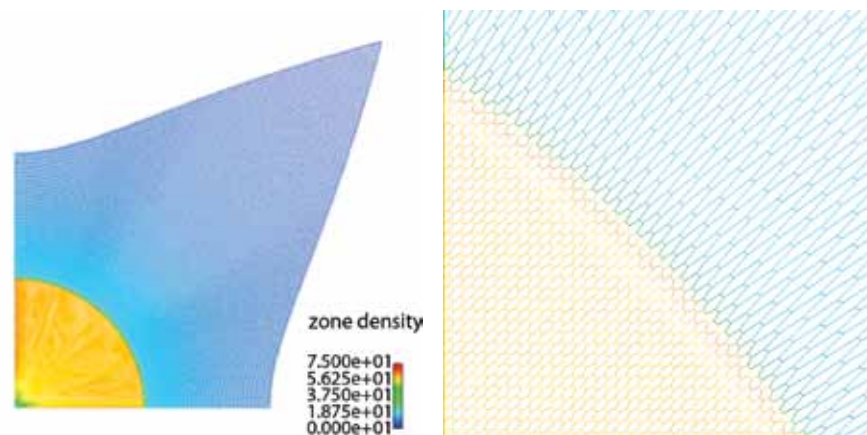


Fig. 2. PENNANT output for Noh problem on a mesh with hexagonal cells.

that a small, self-contained code can capture the basic physics algorithms and data structures of larger multi-physics codes that use unstructured meshes.

Future development plans for PENNANT at LANL include

- additional test problems;
- additional optimizations for serial and multicore versions;
- multi-node version using the Message-Passing Interface (MPI) library to distribute work across the nodes in a cluster;
- GPU versions using other programming models (OpenCL and/or OpenACC);
- and testing on Intel Many-Integrated-Cores (MIC) architecture.

PENNANT will also be made available to other research sites, hardware vendors, and compiler vendors as a research tool to help make unstructured mesh algorithms run more effectively on future hardware platforms and programming models.

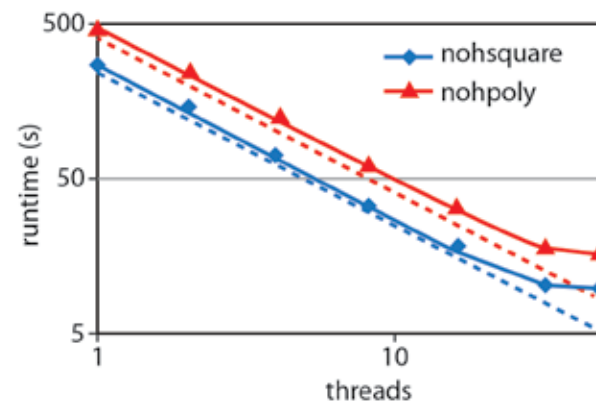


Fig. 3. PENNANT timings running for nohsquare and nohpoly test problems, using up to 48 cores.

- [1] Campbell, J. and M. Shashkov, *J Comput Phys* **172**, 739 (2001).
- [2] Caramana, E.J. et al., *J Comput Phys* **146**, 227 (1998).
- [3] Caramana, E.J. and M.J. Shashkov, *J Comput Phys* **142**, 521 (1998).
- [4] Ferenbaugh, C.R., "A Copomparison of GPU Strategies for Unstructured Mesh Physics," *Concurrency and Computation: Practice and Experience*, in press (2012).
- [5] Ferenbaugh, C.R., "The PENNANT Mini-App: Unstructured Mesh Hydrodynamics for Advanced Architectures," *Proceedings of the Nuclear Explosives Code Developers Conference* (2012).
- [6] Noh, W.F., *J Comput Phys* **72**, 78 (1987).



# Parallel Log Structured File System (PLFS)

**Gary Grider, HPC-DO;**  
**John Bent, EMC Corporation;**  
**Chuck Cranor, Carnegie**  
**Mellon University;**  
**Jun He, New Mexico**  
**Consortium;**  
**Aaron Torres, HPC-3;**  
**Meghan McClelland, HPC-5;**  
**Brett Kettering, HPC-5**

To improve the checkpoint bandwidth of critical applications at LANL, we developed the Parallel Log Structured File System (PLFS)[1]. PLFS is a transformative I/O middleware layer placed within our storage stack. It transforms a concurrently written single shared file into non-shared component pieces. This reorganized I/O has made write size a non-issue and improved checkpoint performance by orders of magnitude, meeting the project's L2 milestone to show increased performance for checkpointing with LANL codes. LANL is working together with EMC under an umbrella Cooperative Research and Development Agreement (CRADA) to further enhance, design, build, test, and deploy PLFS. PLFS has been integrated with multiple types of storage systems, including cloud storage, and has shown improvements in file storage sizes and metadata rates.

A single shared file written concurrently by a large number of processes and non-optimal write sizes is a challenging workload for the current generation of parallel file systems. Despite severe bandwidth limitations, single shared file checkpointing continues to be used heavily in the high-performance computing (HPC) community because it is convenient from a user standpoint. PLFS transparently rearranges these challenging user I/O patterns into patterns optimized for the parallel files system. This reorganized I/O has made write size a non-issue and improved checkpoint performance by orders of magnitude measured to be as much as 150× with improvements in write, read, and metadata performance of our I/O workloads.

As part of the LANL umbrella CRADA with EMC Corporation, LANL and EMC are working to design, enhance, build, test, and deploy PLFS. This collaboration is focused on support for the DOE's exascale initiative and other data-intensive programs and is aimed at boosting HPC capability to ensure efficient resource utilization on the largest supercomputers in the world.

Using PLFS for data placement, LANL and EMC have prototyped a “burst buffer” storage stack. Economic projections into the exascale era dictate that the storage stack be re-engineered to incorporate an intermediary layer between the compute nodes and the disk-based scratch storage space. Within this intermediate layer, applications may conduct in-transit analysis and other data-mining operations, allowing for efficiencies in results collection and file archiving. LANL and EMC have shown overall time to problem solution reduced by 23%, with I/O to the burst buffer 3.5–4× faster than directly to Lustre, an open-source parallel file system. A major modification has been the addition of an I/O Forwarding

Scalability Layer (I/OFSL), a transport mechanism for a scalable I/O layer that scalability models suggest will be required at exascale.

An increasing number of clusters are being configured for data analytics using the Apache Hadoop open-source version of the Google Internet services tool suite (Google File System, BigTable, etc.) [2,3]. Because HPC and Internet services analytics can both be big data and big compute applications families, it is desirable to be able to mix and match applications on cluster infrastructure. Prior research has demonstrated that Hadoop applications can run efficiently on HPC cluster infrastructure [4]. However, cloud storage systems—such as the Hadoop Distributed Filesystem (HDFS)[5]—are not POSIX-based and do not support multiple concurrent writers to a file. In order to enable the convergence of HPC and cloud computing on the same platform, we adapted the PLFS to enable HPC applications to be able to read and write data from the HDFS cloud storage subsystem. Our enhanced version of PLFS provides HPC applications with the ability to concurrently write from multiple compute nodes into a single file stored in HDFS, thus allowing HPC applications to checkpoint. This work paved the way for PLFS to be able to store to multiple types of backends, increasing its overall versatility and usefulness. Research into data and data log compression capabilities in storage backends is showing very promising results and is planned for a future PLFS release.

Current approaches for metadata handling cannot scale to exaflop supercomputers due to the large overhead of creating and reassembling the metadata. We have developed and evaluated algorithms by which patterns in the PLFS metadata can be discovered and then be used to replace the current metadata [6]. Our evaluation shows that these

patterns reduce the size of the metadata by several orders of magnitude, the performance of writes by up to 40%, and the performance of reads by up to 480%. These improvements can allow current checkpointing models to survive the transition from petascale to exascale. Another feature improvement for PLFS is write buffering—the capability of PLFS to buffer writes, reducing the overhead of making many system calls and performing small writes. Flat file mode reduces the number of underlying files and directories created for N-N workloads, decreasing the overall metadata workload and thereby increasing performance.

PLFS met the project's L2 milestones, demonstrating up to 76% increased performance over I/O directly to the file system with LANL codes. PLFS has been shown to work on real problems for real production simulation applications at LANL that use message passing interface (MPI)/IO for N-1 I/O workloads. We have shown that for a benchmark application configured to produce a maximum write bandwidth workload, PLFS does not adversely affect N-N performance. PLFS shows substantial performance improvement over MPI/IO for N-1 workloads for these production simulation codes. Furthermore, PLFS is a general solution, like MPI/IO, that can be used by any application. PLFS is capable of further improving I/O performance by aggregating multiple file systems into a larger virtual file system when the application I/O load justifies the larger file system capabilities. PLFS is integrated into the LANL production system parallel tools installation process and is available for use on LANL HPC platforms.

PLFS is open source software available from  
<http://sourceforge.net/projects/plfs>.

#### **Special Thanks**

Project contributions by the HPC-5 I/O team, the HPC-3 HPC Tools team, Garth Gibson, EMC Corporation Engineering Team, David Knaak, William Tucker, and Steve Oyanagi.

[1] Bent, J. et al., "PLFS: a Checkpoint Filesystem for Parallel Applications," *Proceedings Conference on High Performance Computing, Networking, Storage and Analysis*, 2009.

[2] Ghemawat, S. et al., *Operating Systems Review* **37**(5), 29 (2003).

[3] Chang, F. et al., "A Distributed Storage System for Structured Data," *USENIX OSDI 2006* (2006).

[4] Ananthanarayanan, R. et al., "Cloud Analytics: Do We Really Need to Reinvent the Storage Stack?" *Proceedings 1st USENIX Workshop on Hot Topics in Cloud Computing (HOTCLOUD '2009)* (2009).

[5] Shvachko, K. et al., "The Hadoop Distributed File System," *MSST 2010* (2010).

[6] He, J. et al., "Discovering Structure in Unstructured Data," *Proceedings 7th Workshop on Parallel Data Storage (PDSW '12)* (2012).

# Yellow to Turquoise Integration (YeTI) Project

**Brett Kettering,**  
**Benjamin McClelland, HPC-5;**  
**Kyle Lamb, HPC-3;**  
**Alex Malin, HPC-DO**

Security and access control considerations have prompted LANL to have two completely separate, but essentially mirrored, unclassified high-performance computing (HPC) networks. These networks, called Yellow and Turquoise, share many common but separate components and infrastructure design, as well as the personnel who maintain and upgrade them. Due to inefficiencies and duplicated costs, the Yellow-to-Turquoise Integration (YeTI) project was created to investigate the merging of these network and HPC resources while addressing the security and access controls that prompted the separate networks in the first place. The YeTI project exceeded its initial investigation goal and has implemented the YeTI model on three production HPC machines and has plans for converting all remaining machines in FY13.

**S**haring expensive resources and combining the Yellow and Turquoise networks is expected to save LANL millions of dollars in personnel and hardware costs while maintaining demanding standards for the protection and separation of data.

The implementation chosen for YeTI allows the most expensive infrastructure, such as the HPC compute resource, network switches, parallel and network file system space, and archive space, to be shared. System front-ends are not shared, so that we can identify the user's protection level and limit the user's access to those data.

Users may only access data for which they have been authorized and have the appropriate need to know. Turquoise users only see the Turquoise level of data and cannot access Yellow data. New levels of data protection, such as a section for individuals collaborating with a private company, can be added with much less effort than fielding a new system.

HPC resources are configured on a per job basis to allow for computing with data appropriate for the user who is launching the job without hardware partitions or enclaves. There is minimal impact on the user experience on YeTI machines. In some cases, especially where a user operates in both the Yellow and Turquoise environments, the user experience is improved.

Yellow users can still copy and move files between the Turquoise and the Yellow and can use their existing test and development frameworks. Cluster-to-desktop visualization capabilities are still functional and access to the Reconfigurable Advanced Visualization Environment (RAVE) is maintained. There are no changes from the user point of view for managing user and group permissions or other data protection mechanisms.

Originally, users on the Yellow network could only access Turquoise-network systems by going through gateway nodes. Binaries had to be built on special Yellow network compile nodes and then pushed through the gateway nodes to the Turquoise network systems. Direct checkout/checkin and automated development processes was not possible with this setup.

With the new YeTI model, we have provided a system front-end in the Yellow network address space that is connected through a carefully controlled and protected link to the system's private out-of-band network. The existence of the front-end in the Yellow network address space gives Yellow-network-based users access to normal Yellow network resources (e.g., TeamForge, other source code repositories, Mercury, etc.) and allows most development, debugging, testing, etc., processes to occur as they do on systems currently fully based in the Yellow network. The private network link gives the Yellow-network-based users access to the Turquoise-network-based resources of the system (e.g., compute nodes, parallel file systems, etc.). Turquoise network users see no difference in how the system is currently accessed and used.

Future work for the YeTI project includes investigating methods for reducing dependence on user settings for file access control and removing the impacts associated with potential privilege escalations.

The YeTI project has increased the likelihood of unclassified collaborations with other agencies, universities, and laboratories while increasing the security of our open network operations. These security measures are accomplished with no significant development of local software or large purchases.



## Phase 1: Make Moonlight “Y” & “T” class compute resource in “T” space, then others

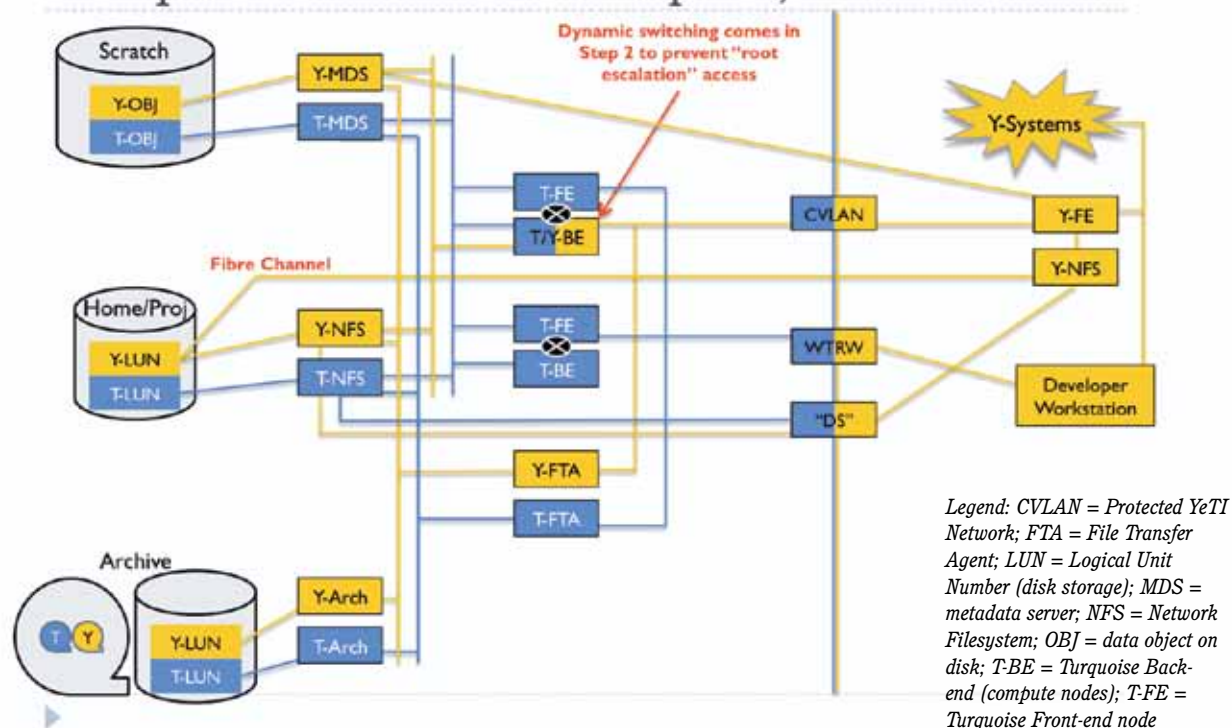


Fig.1. Diagram of the network and cluster modifications for the projects.

# Epidemic Modeling and Facemask Usage

Susan M. Mniszewski,  
CCS-3;  
Sara Y. Del Valle,  
Reid Priedhorsky, D-4;  
James M. Hyman,  
Kyle S. Hickman, Tulane  
University

Evidence from the 2003 severe acute respiratory syndrome (SARS) epidemic and 2009 H1N1 influenza pandemic shows that facemasks can be an effective non-pharmaceutical intervention in minimizing the spread of airborne viruses. Recent studies have shown that the use of facemasks is correlated to an individual's age and gender, where females and older adults are more likely to comply than males or youths. We use the EpiSimS agent-based simulation to model the use of facemasks and quantify their impact on three levels of influenza epidemic. Our results show that facemasks alone have limited impact on the spread of influenza except to the wearer. However, when facemasks are combined with other interventions such as hand sanitizer, they can be more effective. We also observe that monitoring Twitter can be a useful technique to measure compliance. We conclude that educating the public on the effectiveness of masks to increase compliance can reduce morbidity and mortality.

Pharmaceutical interventions such as vaccines and antiviral medication are the best defense for reducing morbidity and mortality during an influenza pandemic. However, the current egg-based vaccine production process can take up to six months for the development and availability of a strain-specific vaccine and antiviral supplies may be limited. Fortunately, alternative strategies, such as non-pharmaceutical interventions, can reduce the spread of influenza until a vaccine becomes available. Facemasks have been used to combat airborne viruses such as the 1918–1919 pandemic influenza, the 2003 SARS outbreak, and the most recent 2009 H1N1 pandemic. These studies indicate that if facemasks are readily available, they may be more cost-effective than other non-pharmaceutical interventions, such as school and/or business closures.

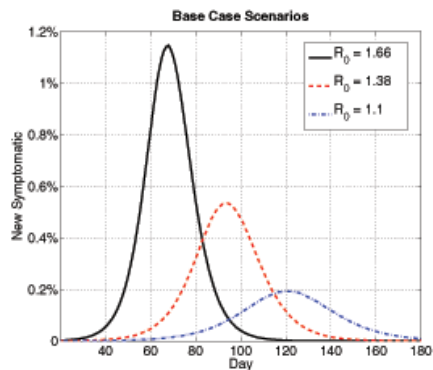


Fig. 1. Base case simulation results for the three different epidemic strengths, showing the percentage of the population that becomes symptomatic per day.

We focus on the use of surgical facemasks and N95 respirators. A surgical facemask is a loose-fitting, disposable device that prevents the release of potential contaminants from users into their immediate environment. They are designed primarily to prevent disease transmission to others, but can also be used to prevent the wearer from becoming infected. If worn properly, a surgical mask can help block large-particle droplets, splashes, sprays, or splatter that may contain germs (viruses and bacteria) and may also help reduce exposure of saliva and respiratory secretions to others. By design, they do not filter or block very small particles in the air that may be transmitted by coughs or sneezes.

A survey paper on demographic determinants of protective behavior [1] showed that compliance in using facemasks is tied to age and gender. They observed that females and older adults were more likely to accept protective behaviors than other population groups. Supporting these ideas, use of facemasks was consistently higher among females than male metro passengers in Mexico City during the 2009

H1N1 pandemic [2]. Limited studies suggest that there is more social stigmatization associated with wearing facemasks in western countries than in Asia. For example, people rarely wear facemasks in public in the US, compared with their use in Japan and China. An article published in 2009 by *New York Times Health* reported that “masks scare people away from one another” resulting in an unintentional social distancing measure or “stay away” factor. During the 2003 SARS outbreak, non-pharmaceutical interventions, where implemented, were seen to follow the epidemic curve [3]. That is, as the perception of SARS increased, more measures were implemented, and as the incidence declined, several measures were relaxed.

Influenza epidemics of varying strengths (high, medium, low) were modeled using EpiSimS [4] [5] to compare the impact of facemask usage on controlling disease spread. These different levels share a similar disease progression. The high-level epidemic is based on the 1918–1919 H1N1 “Spanish flu” outbreak and has high morbidity and mortality, the medium level is based on the 1957–1958 H2N2 Asian flu, and the low level is based on the more recent 2009 novel H1N1 flu. The number of hospitalizations and deaths were extrapolated from the US population during the represented pandemic year to the US synthetic population (based on 2000 census data). The attack rate (percentage of population infected), clinical attack rate (percentage of population symptomatic), hospital rate (hospitalizations out of population), and mortality rate (deaths out of population) are shown for each strength in Table 1. Figure 1 shows each of their respective epidemic curves for the new symptomatic as a function of time.

**Table 1.** Epidemic parameters associated with high, medium, and low strengths of epidemic.

Epidemic Level	Attack Rate (%)	Clinical Attack Rate (%)	Hospital Rate (%)	Mortality Rate (%)
High	40.0	30.0	0.500	0.300
Medium	30.0	19.7	0.250	0.100
Low	20.0	10.0	0.008	0.015

For any intervention, it is important to measure the rate at which it is actually happening. Non-pharmaceutical interventions, such as the wearing of facemasks, presents special problems in this regard because the decision to comply or not comply is an individual one that takes place away from observation by health providers. The social internet system, Twitter, was used to evaluate two conjectures: (1) that the level of facemask wearing follows the disease incidence level, and (2) that analysis of the public Tweet stream is a feasible technique to measure compliance with facemask wearing (and, by implication, other behaviors relevant to infectious disease). To do so, we analyzed Tweets (with a simple keyword-based approach) published globally between September 6, 2009, and May 1, 2010, roughly corresponding to the H1N1 pandemic flu season in the US. Results are shown in Fig. 2. We compare our Twitter mention and observation counts against influenza-like illness (ILI) data published by the Centers for Disease Control (CDC). The correlation is excellent—0.92 for mentions and 0.90 for observations. We expect future efforts to deepen this capability, providing results segmented by locale or demographics.

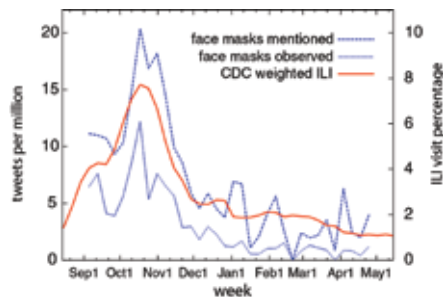


Fig. 2. Comparison of Twitter mention and observation counts against ILI data published by the CDC.

Facemask mitigation strategies were considered for surgical masks and N95 respirators. All scenarios began when 0.01% or 1.0% of the population was symptomatic. Usage was based on age and gender and followed the course of the epidemic. Effectiveness was based on previous testing. Our results show that, in general, facemasks have an impact on reducing the overall incidence and extending the length

of the epidemic. Masks alone reduce the clinical attack rate, on average, by over 10% for the entire population and 50% for the population that wears facemasks. Furthermore, the results are consistent with other studies concluding that the earlier interventions are put in place, the greater the impact they have on reducing morbidity and mortality.

We compare the impact of combining facemasks with hand sanitizers (M&HS) or with social distancing (M&SD). M&HS are assumed to reduce the transmission rate by 50% and M&SD are assumed to reduce the transmission rate by 30%. Figure 3, parts A and C, shows the epidemic curves when M&HS are implemented after 1.0% of the population is symptomatic and M&SD when 0.01% of the population is symptomatic, respectively. In addition to showing the overall dynamics of these two interventions, we show the epidemic curve for individuals who adopted the specified behavior, but who still became infected. Note that although the clinical attack rate was only reduced by 19% and 21% for these two scenarios, the clinical attack rate for M&HS users was only 3.6% or an 81% reduction. Similarly, the clinical attack rate for the M&SD users is 4.7% or a 76% reduction from the base case. Parts B and D show the clinical attack rate for various assumptions of the M&HS and M&SD scenarios and all the different pandemic levels. Not surprisingly, our results show that facemasks are more effective when coupled with other interventions.

We conclude that for mathematical models of infectious diseases to be useful in guiding public health policy, they need to consider the impact of non-pharmaceutical interventions. Facemasks can be a cost-effective intervention when compared to closures; therefore, public health campaigns should focus on increasing compliance. Additionally, measuring the effect of these campaigns should include analysis of social internet systems and other emerging data sources. The results presented here are useful in providing estimates of the effects of non-pharmaceutical interventions on the spread of influenza.

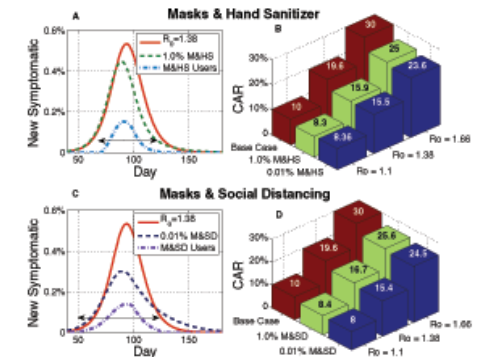


Fig. 3. Results of surgical masks and hand sanitizers (top) and masks and social distancing (bottom).

- [1] Bish, A. and S. Michi, *Br J Health Psychol* **15**, 797 (2010).
- [2] Condon, B.J., and T. Sinha, "Who is that Masked Person: The use of Facemasks on Mexico City Public Transportation during the Influenza A (H1N1) Outbreak," *Health Policy* (2009).
- [3] Pang, X. et al., *JAMA* **290**(24), 3215 (2003).
- [4] Stroud, P. et al., *J Artif Soc Simulat* **10**, 49 (2007).
- [5] Mniszewski, S. et al., "Understanding the Impact of Facemask Usage through Epidemic Simulation of Large Social Networks," *Modelling and Simulation of Complex Social Systems*, Springer, in press (2013).



# Co-Design for Molecular Dynamics: An Exascale Proxy Application

Jamaludin Mohd-Yusof,  
Sriram Swaminarayan, CCS-7;  
Timothy C. Germann, T-1

Roadrunner demonstrated the need to understand the interaction between high-performance hardware and software as early in the development process as possible. One mechanism being used to explore this interaction is the notion of proxy applications that provide a means to test both hardware and software modifications while retaining the essential workload of a real application. Co-designed Molecular Dynamics (CoMD) is one of the proxy “apps” being employed for this purpose by the Exascale Co-Design Center for Materials in Extreme Environments (ExMatEx). We present examples of the options available within CoMD and their effects, as well as potential impacts of future hardware modifications.

**M**olecular dynamics (MD) simulations represent a significant fraction of the DOE workload, as they provide the fidelity required to examine materials’ response to extreme conditions. Examples of interest include nuclear reactor lifetime extension and nuclear stockpile aging. Achieving the fidelity required to simulate these problems at scale will continue to require the largest computational resources available.

High-performance computing (HPC) is currently undergoing a transition period where a variety of new architectures are being explored.

Roadrunner was the first example of a hybrid supercomputer, and required a huge effort to enable codes to run on it [1]. The options

available for future exascale machines include accelerators derived from graphics processing units (GPUs), many-core accelerators such as the Intel Many-Integrated-Core (MIC) architecture, and evolutions from existing CPU architectures from AMD, Intel, and IBM. Each of these hardware choices presents tradeoffs in terms of their relative performance when running scientific computing algorithms.

These algorithms, in turn, represent

varying workloads to the machine, in terms of both intra-node and inter-node requirements.

ExMatEx is one of several efforts to develop a framework in which these hardware-software interactions can be explored. This will enable both the hardware vendors and software developers to co-optimize their products to ensure that future machines are able to provide the performance needed to solve these challenging problems. Part of this strategy uses proxy applications (“proxy apps”) that encapsulate the workload of an actual science application. These simplified applications are more amenable to testing and analysis than existing production applications and are available to external collaborators.

The CoMD proxy app represents the typical workload and use cases of MD simulations of material dynamics. We expect that algorithmic improvements and optimizations will ultimately be incorporated into DOE production MD codes such as LAMMPS, ddcMD, and SPaSM. CoMD represents the fundamental workflow in such simulations from problem setup, equilibration, time integration, analysis/visualization, to checkpoint/restart.

For the development of CoMD, the SPaSM code was chosen as a starting point because the code transformations made to port SPaSM to Roadrunner form a good basis for those needed to optimize an exascale code. In particular, the separation of local work and inter-node communication to reduce latency provide a framework that allows us to concentrate on optimizing the local (intra-node) performance.

MD solves Newton’s laws of motion for individual atoms—for short-range interatomic potentials typical of metals and other neutral, non-ionic systems this requires the evaluation of an interatomic force on each atom arising from all neighboring atoms within a potential-dependent cutoff radius. For our targeted class of materials (metals) that radius typically corresponds to a few interatomic spacings within a lattice, so only 10-100 atoms are within the cutoff. This informs the choice of domain decomposition (spatial) and the use of a link-cell formulation for ordering particles and iterating particle pairs (see Fig. 1). In typical production runs, approximately 95% of the simulation time (aside from checkpoint I/O and in situ analysis/visualization, which vary widely depending on needs) is spent computing the forces on atoms, giving a clear hot-spot for optimization efforts.

In order to explore the tradeoffs discussed earlier, various options are incorporated into CoMD.

- **Execution models.** The base version of CoMD is a simple serial code and provides the reference for all other variants. OpenCL and OpenMP versions are also provided to allow testing on GPUs and multi/many-core processors.

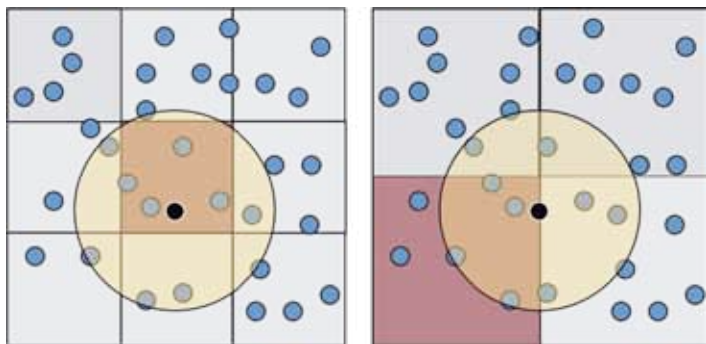


Fig. 1. Sketch showing an example of link-cell data decomposition. The link cells are larger than the cutoff radius for the interatomic potential, so that a search of neighboring cells covers all particles within the cutoff. The tuning parameter box factor is the ratio of the link cell size to the cutoff and can be varied to tune the granularity of the data decomposition. The left sketch shows a box factor of 1, while the right shows the same data decomposed with box factor 1.5.

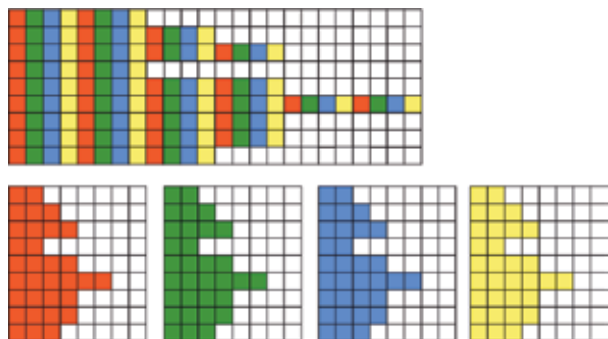


Fig. 2. Sketch of AoS, (top) and SoA, (bottom) data layouts. The different colors represent different data components, for example, the  $x$ ,  $y$ , and  $z$  components of particle position. In the CoMD code, typical AoS data would include the three components of particle position and the mass. Each data layout provides different benefits in terms of data locality and vectorizability of operations, among other factors.

• **Data Layout.** The natural way to think of particle data is as a structure—for example, the position, velocity, and force each have three components  $[x, y, z]$ . In memory, we can group these three components together to form an Array-of-Structures (AoS, Fig. 2a). Alternatively, we could

separate the  $x$ -component data (for all the particles) in contiguous arrays, forming a Structure-of-Arrays (SoA, Fig. 2b). The base version of CoMD utilizes an AoS data layout that mimics that in SPaSM. The OpenCL version has both AoS and SoA data layouts to evaluate the relative performance on a variety of architectures. An example of the importance of data layout is shown in Fig. 3.

• **Potential Representation.** For metals, Embedded Atom Method (EAM) tabulated data is traditionally used to represent the interatomic potential. These tables are too large to fit into the shared cache of many newer architectures, so we provide the option to use polynomial approximations which require a small (user-selectable) number of coefficients but increase arithmetic intensity.

• **Data Decomposition.** Depending on the arithmetic intensity of the algorithm, the performance of the code may be determined by the ability to access memory efficiently. We provide a “box factor,” the ratio of link cell size to the cutoff radius (see Fig. 1), as a tuning parameter. This makes more efficient use of loads from main memory and also reduces the frequency of re-sorting particles between link cells, which can itself be a bottleneck in some situations.

Although it is helpful to measure the performance differences between different implementations on current hardware, we also need to understand the effects that hardware changes may have on performance. Aspen [2] is a framework for the analytical modeling of exascale

applications and architectures that is based on a domain-specific language. The Aspen developers at ORNL constructed a model of CoMD and used it to predict the effect of varying cache size on a GPU, in this case an Nvidia Tesla M 2090, as shown in Fig. 4.

A significant component of the co-design loop is to engage vendors so that changes to future hardware can be incorporated when feasible. CoMD is being used by NVIDIA, AMD, IBM and Intel as a representative workload to study the performance impacts of various design tradeoffs in future processors.

CoMD is one of a suite of proxy apps being developed as part of the ExMatEx project. It provides a variety of options, with more under development, to allow hardware and software developers to understand the tradeoffs needed for future exascale applications. Additional information can be found at <https://github.com/exmatex>.

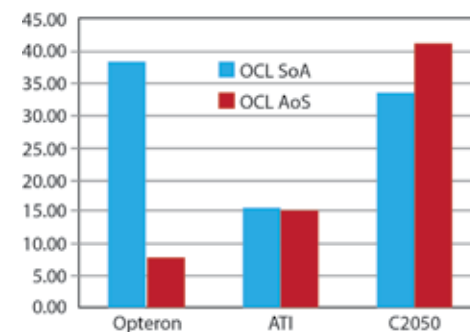


Fig. 3. An example of the effect of data layout choices on application performance. The same OpenCL code is run on an AMD Opteron CPU, ATI Cypress GPU and Nvidia C2050 GPU, using AoS or SoA data layouts. The Opteron performs poorly on the AoS layout. The GPUs show less sensitivity, but differing trends with respect to data layout. The differences may also be attributable to the OpenCL compiler provided by each vendor, showing the importance of the entire hardware/software ecosystem on performance.

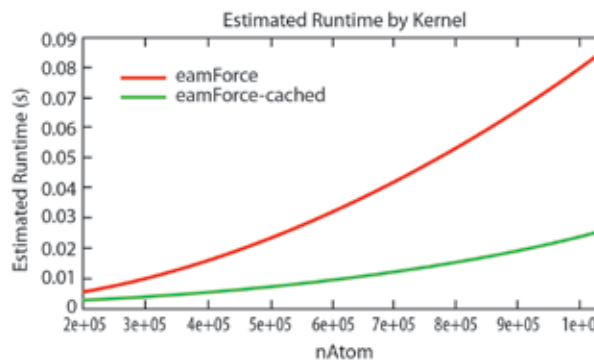


Fig. 4. Predicted performance improvement on an Nvidia M2090 GPU for EAM table lookup, if the cache expanded to hold all the table data. Such predictions can provide important insight for vendors when designing future architectures, as well as for software developers seeking to maximally exploit a given architecture.

[1] First Science at the Petascale: Results from the Roadrunner Supercomputer, October 2010. LA-UR-10-06728

[2] Spafford, K. and Vetter, J.S., “Aspen: A domain Specific Language for Performance Modeling,” SC12: ACM/IEEE International Conference for HPC, Networking, Storage, and Analysis, 2012.

# Accelerating Graph Algorithms Using Graphics Processors: Shortest Paths for Planar Graphs

Hristo Djidjev,  
Sunil Thulasidasan, CCS-3;  
Guillaume Chapuis,  
Rumen Andonov, University  
of Rennes, France

We present a new approach to solving the shortest-path problem for planar graphs. This approach exploits the massive on-chip parallelism available in today's Graphics Processing Units (GPU). By using the properties of planarity, we apply a divide-and-conquer approach that enables us to exploit the hundreds of arithmetic units of the GPU simultaneously, resulting in more than an order of magnitude speed-up over the corresponding CPU version. This is part of our larger algorithmic work on finding efficient ways to parallelize unstructured problems, such as those found in complex networks and data mining, on highly parallel processors.

The shortest-path problem is a fundamental computer science problem with applications in diverse areas such as transportation, robotics, network routing, and Very Large Scale Integration (VLSI) design. The problem is to find paths of minimum weight between pairs of nodes in edge-weighted graphs, where the weight of a path  $p$  is defined as the sum of the weights of all edges of  $p$ . The distance between two nodes  $v$  and  $w$  is defined as the minimum cost of a path between  $v$  and  $w$ .

There are two basic versions of the shortest-path problem. In the single-source shortest-path (SSSP) version the goal is to find, given a source node  $s$ , all distances between  $s$  and the other nodes of the graph. In the all-pairs shortest-path (APSP) version, the goal is to compute the distances between all pairs of nodes of the graph. While the SSSP problem can be solved very efficiently in nearly linear time by using Dijkstra's algorithm [1], the APSP problem is much harder computationally. The fastest algorithm for general graphs is Floyd-Warshall's algorithm [1] that runs in  $O(n^3)$  time and works for graphs with arbitrary (including negative) weights. That algorithm has a relatively regular

structure that allows parallel implementations with high speedup. However, the cubic complexity of the algorithm makes it inapplicable to very large graphs.

As part of our work on solving unstructured graph-based problems on fine-grained parallel architectures, we describe a new algorithm for the

APSP problem for planar graphs based on the Floyd-Warshall algorithm. The complexity of our algorithm with respect to the number of nodes is close to quadratic, while its structure is regular enough to allow for an efficient parallel implementation that enables us to exploit the massive on-chip parallelism of GPUs.

Our algorithm uses the Floyd-Warshall algorithm as a sub-routine, which successively re-evaluates the path between nodes  $i$  and  $j$ , by considering the path through vertex  $k$ , for all possible  $k$ . The structure of the algorithm is similar to the one of matrix multiplication, that makes very regular efficient parallel implementation possible. Our algorithm decomposes the graph into  $p$  parts, solves the APSP problem for the sub-graph induced by each part (in parallel, if more than one processor is available), and then uses that information to compute the distances between pairs of arbitrary nodes. The details of the algorithm are given in [2].

The above algorithm was implemented on a GPU using CUDA, nVIDIA's parallel programming framework for the GPU. Modern GPUs are efficient at manipulating structured data like matrices, and their highly parallel architecture (a GPU trades the complicated cache and control logic in a CPU for a large number—often hundreds—of arithmetic units) makes them ideally suited for processing large blocks of data simultaneously (referred to as the SIMD—Single Instruction Multiple Data—paradigm). In order to comply with the GPU paradigm, we define a computational kernel that implements Floyd-Warshall's algorithm and computes the APSP over a sub-matrix of the initial matrix. We then define computation grids, where blocks correspond to sub-matrices that can be computed simultaneously.

Phase 1 of the algorithm consists of computing APSP in self-dependent (in terms of data dependencies) diagonal sub-matrices, using a

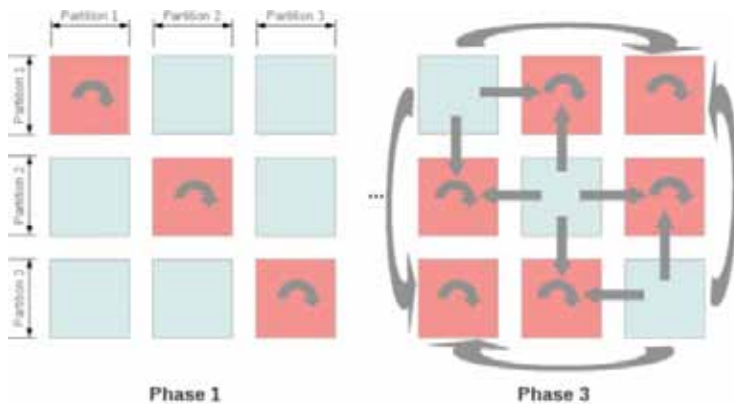
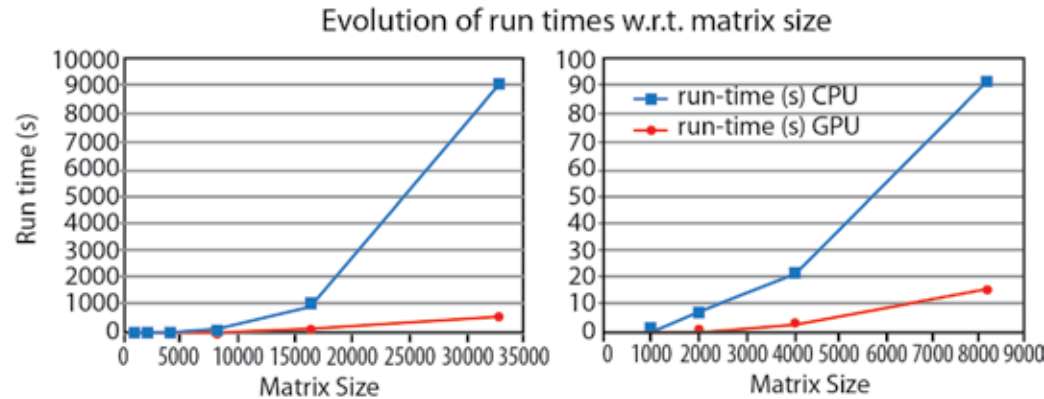


Fig. 1. Block-level data parallelism and data dependencies in the adjacency matrix for phase 1 and phase 3 of the algorithm. Sub-matrices for which computations are required are shown in red. Arrows indicate data dependencies.



Fig. 2. Run times (left) with respect to input matrix size or the CPU and GPU versions. First four data points (right) zoomed in.



block-parallel version of the Floyd-Warshall algorithm. Computations for these diagonal sub-matrices can be run in parallel in different GPU blocks. Phase 2 of the algorithm consists of computing APSP on a sub-graph of the initial graph solely comprised of boundary vertices. For this purpose, we implemented a GPU version of the blocked APSP algorithm described in [5], a variant of the Floyd-Warshall algorithm where computations are divided into groups that can be easily be mapped to GPU blocks. Phase 3 of the algorithm consists of computing APSP in the remaining non-diagonal sub-matrices, using the same parallel Floyd-Warshall algorithm. Since Phases 1 and 3 use the same data patterns we illustrate their data dependencies in Fig. 1.

In order to test the efficiency of the algorithm, we compare two implementations of the partitioned APSP algorithm: (1) a single core CPU implementation of the partitioned all-pairs shortest-path algorithm (referred to as CPU version), and (2) a single GPU implementation of the partitioned all-pairs shortest-path algorithm (later referred to as GPU version).

The CPU version runs on an Intel(R) Xeon(R) CPU X5675 at 3.07GHz. The GPU version runs on an nVidia Tesla m2090 consisting of 512 cores at 1.3 GHz. The benchmark consists of random cost adjacency matrices, representing planar graphs with sizes ranging from 1024 vertices to 32,768

vertices. These graphs were generated using the LEDA graph generator [3]. Figure 2 shows the GPU versus CPU speed-up comparison for progressively larger graphs, with the GPU being up to 14 times faster than the CPU for larger graphs (32-k nodes). For the largest instance, the cost adjacency matrix requires about 4.2 GB of RAM. Instances larger than about 38,000 vertices would require more RAM than currently available on the GPU. One way around the memory size problem is to exploit the spatially constrained nature of paths in real-world graphs that will allow us to consider only a relatively small subset of the original graph [4]. We are also currently extending this work to tackle larger graph instances by using multi-GPU clusters.

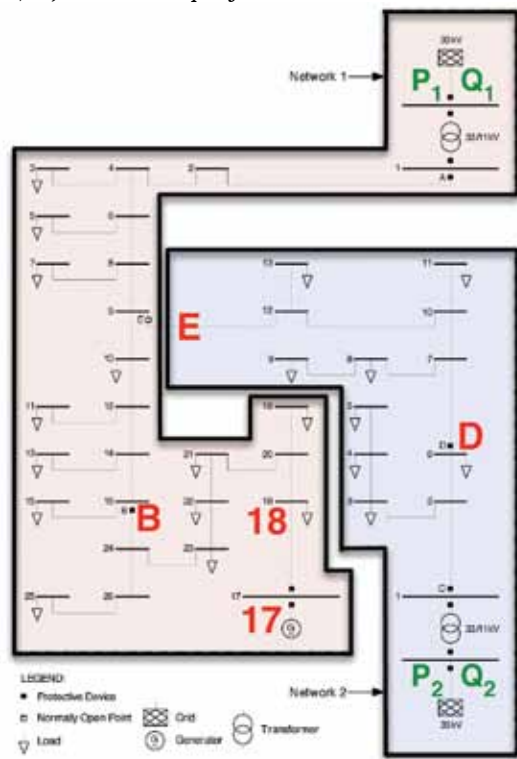
- [1] Cormen, T.H. et al., *Introduction to Algorithms*, 1st edition, MIT Press and McGraw-Hill, ISBN 0-262-03141-8 (1990).
- [2] Djidjev, H.N. et al., "On Solving Shortest Path Problems for Planar Graphs using Graphics Processors," LA-UR-12-25700 (2012).
- [3] Mehlhorn, K. et al., *LEDA: A Platform for Combinatorial and Geometric Computing*, Cambridge University Press, (1999).
- [4] Thulasidasan, S., "Heuristic Acceleration of Routing in Transportation Simulations Using GPUs," *Proceedings 4th International Conference on Simulation Tools and Technologies (SimuTools)* (2011).
- [5] Venkatraman, G. et al., *J Exp Algorithmics*, 8, 2.2 (2003).

# Network Uncertainty on Electric Power Grids

Russell Bent, D-4;  
Earl Lawrence,  
Scott Vander Wiel, CCS-6

Electric grid operators decide how much power each of their generators should produce based on an assumed network state. The state can change as safety devices trigger or lines go down, but these events may not be observed directly so estimates are needed from a limited number of flow measurements at generators and transformers. Demands at internal nodes are usually not measured but can be assigned probability distributions. We estimate the present state of a network and its uncertainty by posterior probabilities that each state in a bank of alternatives could have produced the observed measurements. Randomly drawn demands are repeatedly run through the bank of network simulators to produce statistical distributions of power flows. Flow measurements at any point in time can be used to calculate state probabilities across the bank of models. Ambiguity about the current state suggests optimizing power generation with explicit hedging against unfavorable possibilities.

Fig. 1. A small test network with measurements  $P_1, Q_1, P_2, Q_2$  at two nodes and five contingencies (B, E, D, 17, 18) that alter the topological state.



Infrastructure networks have a prominent role in national and global security because we increasingly depend on them to deliver commodities such as information and energy. However, they are difficult to monitor because of their complexity. We are developing methods for inference about unobserved parts of a network (links, nodes, flows, attributes) using science-based simulations to match measurements on observable components. Our focus is on electric power.

Modern power grid operations (optimizing generation, monitoring exchanges, modifying safety devices, restoring damage, and determining criticality) depend heavily on understanding the current state of the grid. Operators typically have very limited real-time observations of voltages, flows, consumption, production, and even topology, so state estimation is a major topic of interest in the power engineering community [1-3].

Traditional state estimation methods [1,3] are based on relatively simple statistical analyses such as least-squares fitting within a greedy topology search. These methods produce a single estimate of the network configuration that matches available measurements within reason. A more recent approach [2] assumes that a bank of models contains all of the important configurations and attempts to directly estimate a single most likely model. Best-fitting state estimates generally work well, but occasional failures have produced disastrous consequences. For example,

root causes of the 2003 northeast blackout included problems with state estimators combined with operator error [4,5]. If uncertainties were integrated into operations, control algorithms could hedge against unfavorable states that are less likely but still plausible.

We are developing statistical methods to identify the current topology of a power network based on partial measurements. The method that follows uses importance sampling to estimate probabilities for each member of a model bank and thereby constrain the set of plausible grid configurations and appropriately represent uncertainty in the current state.

The small network of Fig. 1 from [2] has 39 buses (nodes) within two subnets that are connected to a larger transmission grid at the top and bottom of the figure. Several devices (denoted by squares) can trigger automatically and disconnect parts of the network or link the two subnets together. As in [2], the model bank consists of a set of network configurations representing contingencies that need to be guarded against. Red labels in Fig. 1 indicate the five major contingencies.

The operator measures only the real and reactive power flowing between the transmission grid and each subnet ( $P_s$  and  $Q_s$  in the figure). Information from these measurements should be used to decide how much power to generate at node 17, trading-off the cost to buy power from the grid and the risk of overloading lines in the contingencies where protective devices have altered the nominal distribution topology.

In addition to the few direct measurements of power, the operator can use probability distributions of historical loads for each internal node and a physics-based alternating current (AC) solver that computes

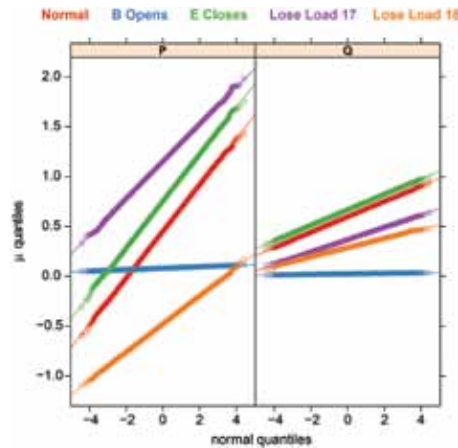
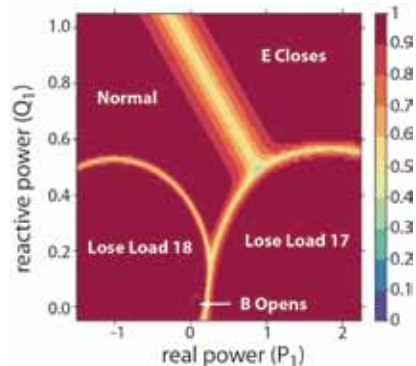


Fig. 2. Normal probability plots of real ( $P$ ) and reactive ( $Q$ ) flows corresponding to  $P_1$  and  $Q_1$  in Fig. 1.

Fig. 3. Classification regions for five contingencies using just measurements  $P_1$  and  $Q_1$ . Color encodes the maximum contingency probability with one being red. Regions of orange, yellow, and green indicate ambiguity in the topological state of the network.



power flowing over lines, provided the interconnection topology and all loads and generation amounts are known.

This setup is small but exemplary of typical power grid state estimation in that only a small fraction of relevant quantities are measured directly while many others can only be given probability distributions. The physical constraints of electrical power flow provide further information relevant to the decisions that need to be made.

Given a measurement ( $P_1, Q_1, P_2, Q_2$ ), we use standard importance sampling to

efficiently calculate state probabilities. The method requires pre-computing, for each enumerated contingency, the mean and covariance of flows at measurement nodes using loads chosen from historical probability distributions. The AC power solver converts the random loads into random flows for each of the contingencies in the model bank. The resulting flows should be transformed to produce approximate Gaussian quantities for ease of implementing the importance sampler. Figure 2 shows normal probability plots of real ( $P$ ) and reactive ( $Q$ ) flows corresponding to  $P_1$  and  $Q_1$  in Fig. 1 when loads at internal nodes have standard deviations that are 20% of historical average values.

Linearity in the plots shows that the flows are Gaussian with means and covariances that depend on the network topology. Measured flows are modeled with 1% standard deviation from the actual values. This completes the off-line preparation for state estimation by importance sampling.

When a measurement ( $P_1, Q_1, P_2, Q_2$ ) is obtained, the importance sampler draws many candidates for the corresponding actual flows, using a standard deviation of 2% around the measured values to cover the full range of possible measurement errors. Estimates of the probabilities for each contingency in the model

bank are obtained through standard importance sampling formulae [6] that use weighted averaging to combine the candidate flows with the Gaussian distributions of Fig. 2.

Importance sampling is fast and accurate for this problem and can easily produce updates of the state probabilities at one-second intervals. Figure 3 shows classification regions for five contingencies using just measurements  $P_1$  and  $Q_1$ . Color encodes the maximum contingency probability with one being red. Regions of orange, yellow, and green indicate ambiguity in the topological state of the network. When measurements fall in the ambiguous regions, decisions such as optimal generation should hedge against state uncertainty. In these cases, the best single estimate of the state is not an adequate description of inherent ambiguities in the partially observed network.

We have applied importance sampling to model bank state estimation for the IEEE Reliability Test System [7], a network of 72 buses and 125 lines, with measurements on five lines that connect three subnets. The model bank consisted of 125 contingencies, one for each of the lines going down. In this case, the measurements generally leave large ambiguities about the network state. The research opportunity is to determine whether a few additional measurement points could resolve much of the uncertainty.

Future research objectives are to optimize power generation under state uncertainty and to incorporate probabilistic risk of overloading lines due to states that are plausible but unfavorable for efficient power distribution. This objective requires quantifying uncertainty of all unobserved flows on the grid. We are exploring the use of Gaussian process emulators to support these inferences with a manageable number of runs of the AC power flow solver.

- [1] Clements, K.A. and P.W. Davis, *IEEE Trans Power Syst* **3**, 1748 (1988).
- [2] Singh, R. et al., *IEEE Trans Power Syst* **25**, 1329 (2010).
- [3] Wu, F.F. and W.E. Liu, *IEEE Trans Power Syst* **4**, 176 (1989).
- [4] Anderson, G. et al., *IEEE Trans Power Syst* **20**, 1922 (2005).
- [5] U.S. Canada Power System Outage Task Force, "Final Report on the August 14, 2003 Blackout in the United States and Canada," technical report, DOE (2004).
- [6] Ripley, B.D., *Stochastic Simulation*, Wiley, New York (1987).
- [7] Reliability Test System Task Force of the Application of Probability Methods Subcommittee, *IEEE Trans Power Syst* **14**, 1010 (1999).



# A Computational Social Model of the Education System

**Stephan Eidenbenz,**  
**Patrick Kelly, CCS-3;**  
**Benjamin Sims,**  
**Joanne Wendelberger, CCS-6**

The question of how to improve the US education system, particularly in science, technology, engineering, and mathematics (STEM) fields, is an important policy issue. Yet there is little consensus on what policy interventions actually work: Smaller classes? Placing top teachers in failing schools? Tracking students by ability level? Education researchers have examined many interventions through rigorous statistical analysis and modeling [1,2]. However, these efforts are often complicated by the difficulty of isolating the effects of a particular intervention from a wide array of other variables—a common problem with assessing “social experiments” in real-world systems. Computational social models represent an alternative way of carrying out such social experiments. Although they cannot capture all the dynamics of a complex social system, such models can incorporate the most important factors. These factors can then be manipulated and monitored without the difficulties associated with doing so in real-world settings. These virtual experiments, so long as their limitations are understood, can be a valuable tool for assessing the potential impacts of policy changes.

**T**his year, LANL completed the initial phase of an exciting collaborative research effort to construct a computational social model of the US education system. Funded by the National Science Foundation and led by the University of California, Santa Cruz, this project also involves education experts from the University of Texas at Austin, the National Laboratory for Education Transformation, the San Jose Unified School District, and the education evaluation firm Gargani and Company. The goal was to have these domain experts interact with LANL statisticians and computational modelers throughout the modeling process to ensure that the final result is not only computationally rigorous, but also relevant to the knowledge and concerns of the education community. This modeling effort builds on LANL’s past success in modeling human behavior in socio-technical systems such as transportation and communication infrastructure [3].

The project takes a three-pronged approach involving conceptual model development, computational model development, and statistical analysis [4].

Conceptual model development is a process in which subject matter experts work with model developers to lay out the key variables that are relevant to a computational model and some of the relationships between the variables. In the current project, this initially took the form of informal conversations among the collaborators, but eventually involved a formalized knowledge elicitation process. A wide range of possible variables was considered, but in the end was cut down to a basic set of entities and properties shown in Fig. 1.

Computational model development is the process of implementing the conceptual model in an agent-based modeling framework. The progress of an individual student through the model in a particular subject area (say, math) is shown in Fig. 2. In order to run the model, a school system is specified consisting of schools, students, teachers, courses, classrooms, and individual classes. The model then assigns teachers and students to individual classes and progresses from one year to the next. Say student X has a medium ability to learn math. Each year, they are assigned a math test score based on their ability, their teacher’s ability, the average ability of the class, and their math score from the previous year (if any). As a result, Student X struggles in Grade 1, but the influence of teachers and peers enables him/her to improve to a high math score by Grade 4.

Statistical data analysis is used to characterize San Jose Unified School District data in order to generate realistic inputs to the agent-based model. For this initial effort, we identified a set of students who could be tracked continuously from Grade 2 to Grade 8 and used Grade 2 test scores to approximate their initial learning ability. We also generated teacher effectiveness scores based on the average improvement in the scores of their students over their scores in the previous year. Based on these inputs, the model was then run for grades 3–8. The model was run numerous times with different weights assigned to the model variables. For each model run, the distribution of predicted student math scores in Grade 8 was compared to the actual distribution of Grade 8 scores in order to identify which set of weights generated the most realistic results. While still in its early stages, this modeling collaboration has already generated significant interest in the education community, and is

likely to continue. Future efforts will focus on expanding the number of variables in the model, refining the relationships between variables, and using statistical methods to generate realistic synthetic populations that can be tailored to address specific research or policy questions [5].

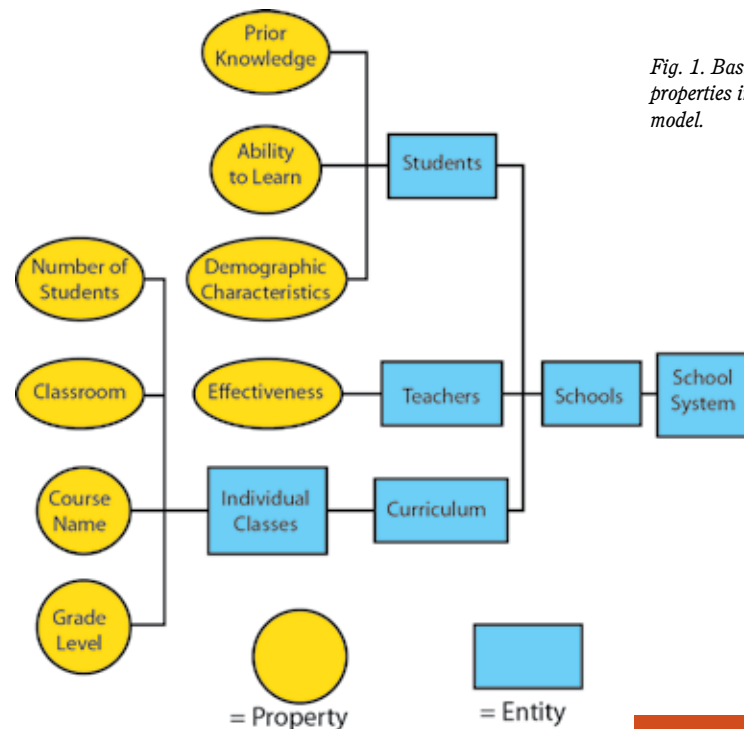


Fig. 1. Basic set of entities and properties in the conceptual model.

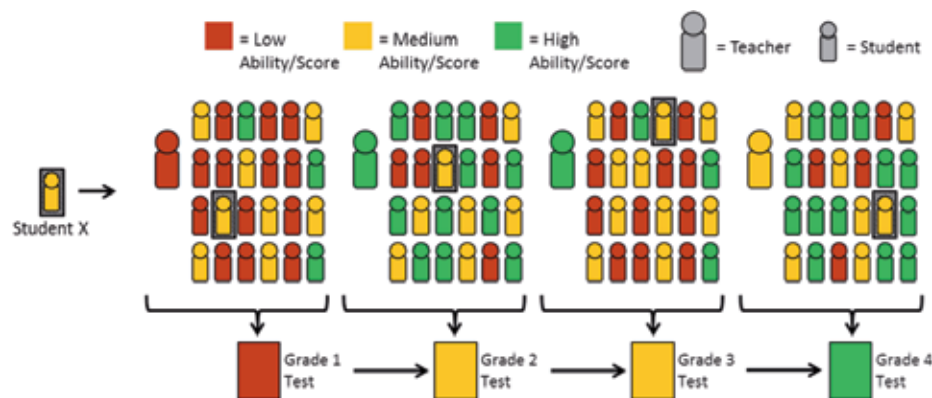


Fig. 2. Student X's progress through the computational model.

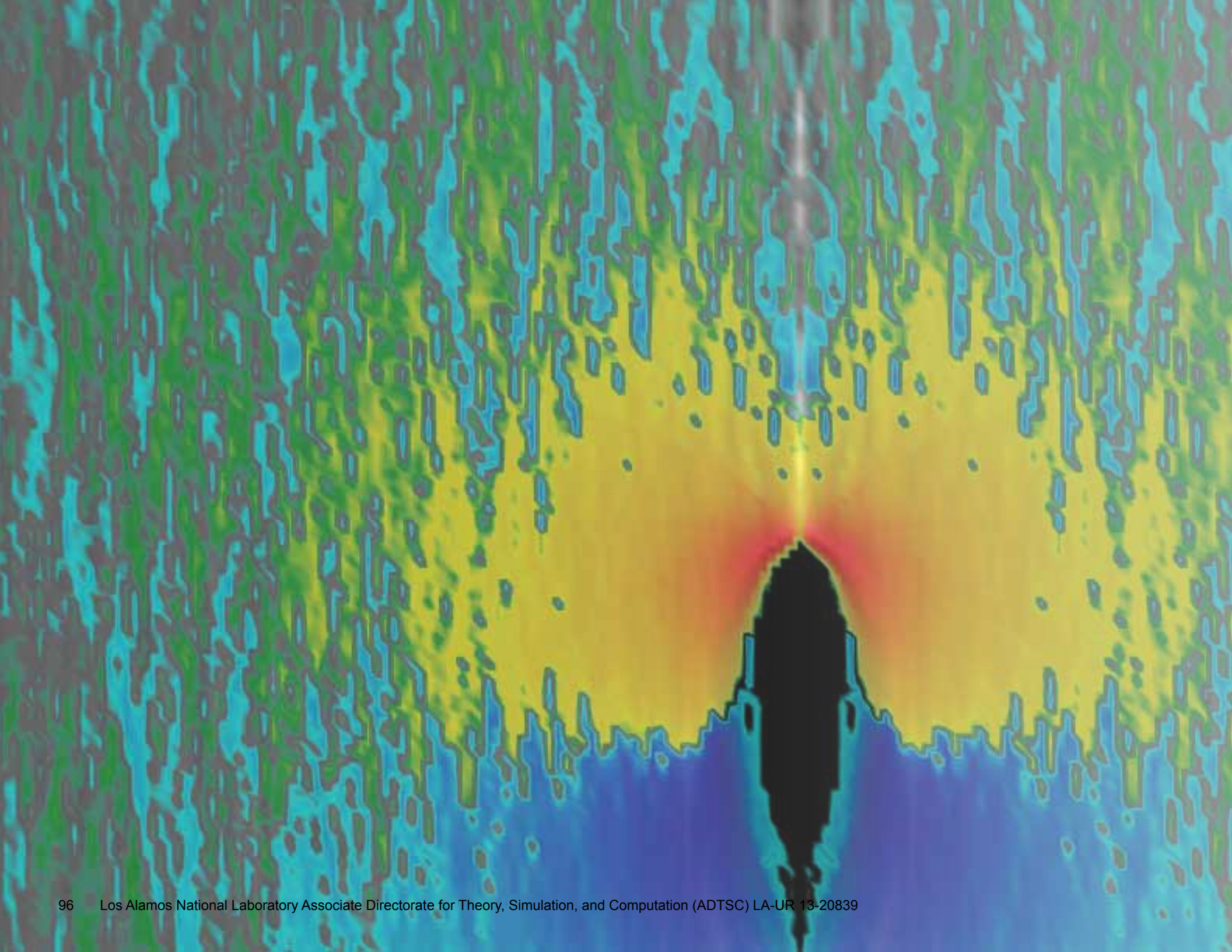
[1] Oppendakker, M. et al., *Sch Effect Sch Improv* **13**, 399 (2002).

[2] Young, D.J. et al., *J Educ Res* **89**, 272 (1996).

[3] Galli E. et al., "ActivitySim: Large-Scale Agent-Based Activity Generation for Infrastructure Simulation," *Proceedings 2009 Spring Simulation Multiconference*, (2009).

[4] Sargent, R.G., *Proceedings 2005 Winter Simulation Conference*, **130** (2005).

[5] Beckman, R.J. et al., *Transport Res Pol Pract* **30**, 415 (1996).





## Atomic, Nuclear, and High-Energy Physics

In this section, we highlight examples of recent research that is representative of the work done in the ADTSC Directorate in support of the mission of LANL and the DOE. LANL has developed a major capability in this area, and the articles in this section exemplify the breadth of our work.

This section includes discussions on increasing the repetition rate in various nuclear magnetic resonance applications, understanding the source of matter-antimatter asymmetry in the universe, calculating local-thermodynamic-equilibrium light-element opacities, using quantum Monte Carlo simulations to investigate the role of the three-neutron force in neutron matter, applying molecular dynamics to the development of stopping models in plasmas, and creating new and better databases of nuclear masses and decay half-lives.

# Radiation Damping for Speeding Up NMR Applications

Gennady P. Berman, T-4;  
Michelle A. Espy,  
Petr L. Volegov, P-21;  
Vyacheslav N. Gorshkov,  
National Technical University  
of Ukraine,  
Vladimir I. Tsifrinovich,  
Polytechnic Institute of New  
York University

We address a problem of low repetition rate in nuclear magnetic resonance (NMR) applications. We suggest using a dynamical regime in the radiation damping that allows returning the nuclear magnetization to its equilibrium state during a time interval that is negligible compared to the spin relaxation time. We show theoretically and numerically that the radiation damping in the spin echo technique can be as effective as application of special radio-frequency (rf)-pulses. We obtain an analytical estimate for optimal damping that is consistent with our numerical simulations.

The problem of low repetition rate is one of the most important in many NMR applications. There are two main approaches for controlling the nuclear magnetic relaxation. The first is the application of an additional rf-pulse in the process of the echo formation. This pulse can return the nuclear magnetization to its equilibrium position. However, in order to implement this method the phase of the rf-pulse must be accurately adjusted relative to the phase of the transversal magnetization. The second method relies on the radiation field created by the resonant circuit interacting with the nuclear magnetization. In this method, the phase of the rf field produced by the circuit is automatically adjusted relative to the transversal magnetization. However, the effect of the radiation field is normally small because the radiation field is much smaller than the rf field. In [1], we demonstrated that the radiation damping can be as effective as the application of an additional rf-pulse. By increasing the duration of the spin echo and optimizing the resonant circuit parameters, one can restore the equilibrium position of the nuclear magnetization during the time of the echo formation, which is negligible compared with the spin relaxation time.

Our consideration in [1] was done in the “dynamical” regime, when the characteristic time of the relaxation in the resonant circuit is much larger than the spin echo time. Our approach is valid for both weak and strong magnetic fields. We obtained an analytical estimate for the optimal conditions for radiation damping that is consistent with our numerical simulations. It is necessary to emphasize that in the dynamical regime considered in [1], the recovery of the longitudinal nuclear magnetization does not depend on the quality factor of the circuit because the recovery time is small compared to the time constant of the resonant circuit. This situation has not been studied in earlier

publications on radiation damping. Also, we should note that from a theoretical point of view the fast radiation damping (FRD) obtained in [1] may not have an advantage compared to the application of an additional rf-pulse. Neither method interferes with signal processing methods for improving the signal-to-noise ratio. The optimal choice of a technique for speeding up the relaxation depends on the concrete experimental conditions (Fig. 1). Coil “0” produces a permanent non-uniform magnetic field,  $B_0$ , in the positive z-direction. Coil “1x” produces an oscillating field  $B_{1x}$  (rf-pulses) along the x-axis. It is also used for NMR detection (with weak radiation damping). Coil “1y” produces an oscillating field  $B_{1y}$  (rf-pulses) along the y-axis. It also can be used for NMR detection (with weak radiation damping). Coil 2 produces the oscillating field,  $B_2$ , along the x-axis, which causes radiation damping. Note that we use two coils,  $1_x$  and  $1_y$ , in order to produce circularly polarized rf-pulses. This is especially important if one is going to use an ultra-low permanent magnetic field,  $B_0$ . In the opposite case of high permanent magnetic field, instead of the circular polarized rf-pulses one can safely apply linearly polarized rf-pulses. In this case one of the coils ( $1_x$  or  $1_y$ ) can be removed.

We propose the following scheme:

- At  $t=0$  a rectangular rf  $\pi/2$ -pulse of duration, say,  $10 \mu\text{s}$  is applied to the sample. At time  $t=1,010 \mu\text{s}$ , a rectangular rf  $\pi$ -pulse of duration of  $10 \mu\text{s}$  is applied to the sample. At time  $t=3,020 \mu\text{s}$ , a rectangular rf-pulse of duration  $10 \mu\text{s}$  is again applied to the sample.
- After the first pulse one observes a decaying signal of magnetic induction. Between the second and the third pulses one observes a spin echo. After the third pulse one again observes a spin echo (see Fig. 2).

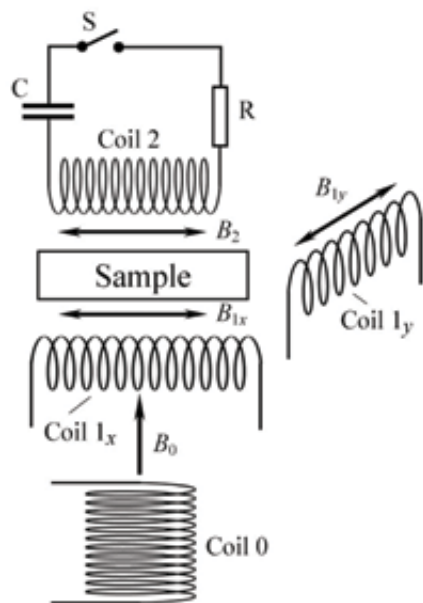


Fig. 1. Suggested setup.

- Coil 2 is a part of an LCRS-circuit, (L stands for inductance; S stands for a switch) with capacitance, C, and resistance, R. The switch, S, is closed (turned on) at  $t=3,030 \mu\text{s}$ , that is, after the third pulse. Before the third pulse, coil 2 does not influence the spin system. During the duration of the second spin echo, the radiation damping due to coil 2 drives the nuclear magnetization toward its equilibrium state.

Finally, in order to increase the repetition rate in various NMR applications we suggested the regime of FRD in the spin echo technique. We have shown that FRD can effectively restore the longitudinal nuclear magnetization during the time of the spin echo formation, which is small compared to the time constant of the resonant circuit and the spin relaxation time. We have obtained an estimate for the optimal choice of the resonant circuit parameters and confirmed it in our numerical experiments.

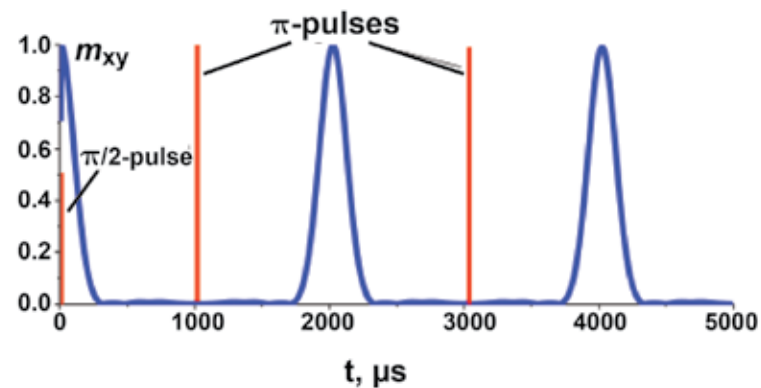


Fig. 2. Suggested spin echo sequence.



# Flavored Baryogenesis

Vincenzo Cirigliano,  
Christopher Lee, T-2

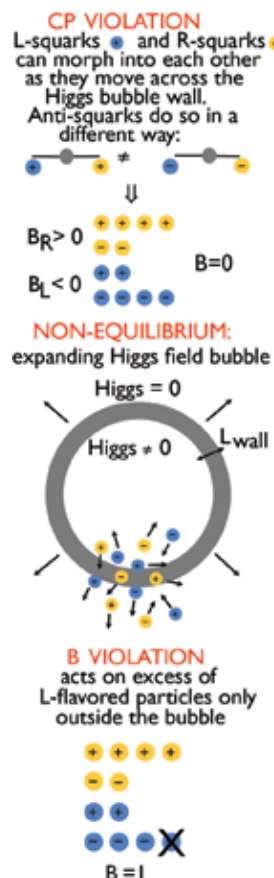


Fig. 1. Stages of “flavored” electroweak baryogenesis: (1) Bubbles of non-zero Higgs field (similar to vapor bubbles in boiling water) form and expand; (2) CP-violating propagation across the bubble generates non-zero L and R baryon densities, with  $B_L + B_R = 0$ ; (3) B violation destroys  $B_L$ , making net  $B \neq 0$ ; (4) The asymmetry is captured by making net  $B \neq 0$ . Finally the asymmetry is captured by the expanding bubble that eventually fills the universe.

Understanding the origin of the matter-antimatter asymmetry of the Universe is one of the greatest challenges at the interface of cosmology, particle physics, and nuclear physics. One of the most attractive mechanisms of “baryogenesis” (from the Greek “generation of matter”) involves particles whose existence can be probed with laboratory experiments. Conclusive tests of this scenario require, beside experiments, robust theoretical calculations of the non-equilibrium quantum transport of particles in the early Universe, which are currently lacking. We have focused on developing the tools that allow calculations of the matter-antimatter asymmetry in a broad class of models and we have identified a new source for the matter-antimatter asymmetry, originating in the quantum mechanical mixing of particles carrying different “flavor” quantum numbers.

To a very good approximation, the laws of physics do not distinguish between matter and antimatter. The Universe would look almost exactly the same if every particle were replaced by its antiparticle. Yet the observed Universe is made up of matter and not antimatter. In order to produce the observed asymmetry, there must have been a tiny imbalance of one part in ten billion between matter and antimatter in the primordial hot plasma one microsecond after the big bang. As the Universe evolved and cooled down, most of the antimatter annihilated with matter, leaving behind the small initial excess of matter, which is still enough to make up the observed stars and galaxies. Understanding how the observed imbalance has developed during cosmic evolution is one of the great open questions at the interface of cosmology, particle physics, and nuclear physics.

A number of explicit baryogenesis mechanisms have been proposed. All mechanisms involve three basic ingredients, first identified by Sakharov in 1967 [1]: (1) violation of conservation of baryon number (= “matter” number), which distinguishes protons from antiprotons; (2) violation of CP symmetry (C= charge conjugation, i.e., interchange of particles and antiparticles; P= mirror reflection); and (3) departure from thermal equilibrium.

We have focused on the so-called electroweak baryogenesis scenario [2], in which the matter-antimatter asymmetry develops during the electroweak phase transition, at which the Higgs field acquires a vacuum expectation value that gives mass to all particles that couple to it. This happened when the temperature of the early universe was about 100 GeV (about 100 times the rest energy of the proton), an energy scale accessible in laboratory experiments—directly at high energy colliders and indirectly via precision low-energy measurements, such as the

search for the neutron electric dipole moment at the ORNL Spallation Neutron Source, led by nuclear physicists from LANL and ORNL. The fact that we can probe experimentally the relevant degrees of freedom implies that in any given extension of the Standard Model we can test whether electroweak baryogenesis works or not. Currently, the main obstacle to such a test stems from large theoretical uncertainties (several orders of magnitude) arising from uncontrolled approximations in the treatment of particle transport in the early universe. We have provided the framework to perform robust first-principles calculations based on a controlled power-counting [3], thus removing the uncontrolled part of the previous treatments. In addition, as we discuss in greater detail below, for a broad class of models we have identified a new dominant source of matter-antimatter asymmetry [4].

We have developed a new formalism and a new physical picture to describe how CP asymmetries (i.e., differences in particle and antiparticle propagation) arise in electroweak baryogenesis. The key observation is that in most extensions of the Standard Model, particles that carry baryon number come in different flavors, that is, feel different interactions or forces. The simplest example of this is provided by left (L) and right (R) “squarks” in supersymmetric models. We have realized that CP asymmetries can arise through coherent flavor oscillations (i.e., morphing of one flavor into another) of all the species that couple to the Higgs particle at the electroweak phase transition. We have then formulated the appropriate transport equations that describe the physics of flavor oscillations and at the same time take into account collisions of particles in the hot environment of the early universe. We have tested the new ideas and formalism within a simplified model that involves only particles of spin zero (such as squarks), providing both numerical solutions and analytical insight. This has to be considered the first step

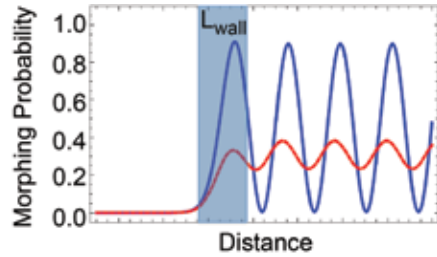


Fig. 2. CP violation at work: As squarks fly across the bubble, the probability for an L-squark to morph into an R-squark (blue line) differs from the probability for an L-antisquark to morph into an R-antisquark (red line).

towards a new understanding of electroweak baryogenesis.

The physics of this mechanism is illustrated in Fig. 1. At the electroweak phase transition bubbles of “broken electroweak phase” (similar to vapor bubbles in boiling water) nucleate and expand. The interesting dynamics happen near the bubble

boundaries, where on a characteristic length scale  $L_{\text{wall}}$  the Higgs field goes from zero to a finite value. As the Higgs field changes, so do the masses and mixing of all particle species that couple to the Higgs, such as L- and R-squarks. If CP symmetry is violated, then L-squarks morph into R-squarks differently than their corresponding antiparticles, as illustrated in Fig. 2. So starting from a CP-conserving equilibrium initial state, a CP asymmetry can arise through coherent oscillations, when the particles cross the boundary between broken and unbroken electroweak phase. In the unbroken phase, there are additional high-temperature processes that then convert this CP asymmetry into the baryon asymmetry that becomes the visible matter in the Universe today.

Within this simple model, we have shown how to derive from first principles, using non-equilibrium field theory, the quantum mechanical Boltzmann equations that describe the dynamics of CP-violating flavor oscillations and collisions in the early Universe. Our derivation can be generalized to any model because it is based on a new power counting based on the hierarchy of relevant length scales: the de Broglie wavelength of the particles, collision mean free path, and flavor oscillation length.

Using numerical methods, we have solved the transport equations exactly without ansatz for the functional form of the density matrices (as far as we know this is the first full solution of quantum kinetic equations in a spatially non-homogeneous system). We have demonstrated the existence

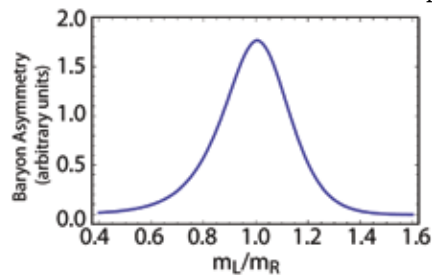


Fig. 3. Matter-antimatter asymmetry versus the ratio of L- to R-squark masses. The Large Hadron Collider might discover supersymmetry and measure  $m_L/m_R$ .

of a resonant enhancement in matter production when the flavor oscillation length is comparable to the thickness of the bubble wall. The resonance is determined by the relative value of the mass parameters of the mixing particles in the underlying theory (see Fig. 3). We have determined how the width of the resonance depends on the parameters of the underlying theory, and have shown that the resonant enhancement is relevant in significant regions of the supersymmetric parameter space.

We have also investigated the impact of collisions in various regimes that correspond to different mixing particles in realistic extensions of the Standard Model. As expected, larger couplings (more frequent collisions) reduce the generated asymmetry near the bubble and lead to shorter density tails in front of the bubble wall, thus suppressing the final matter-antimatter asymmetry.

Finally, we have compared our results with the existing “state-of-the-art” treatments of this problem by Konstandin, Prokopec, Schmidt, and Seco (KPSS) [5]. Comparison of our exact solution with the KPSS approximation scheme reveals dramatic failures of the latter, as illustrated in Fig. 4. We conclude that: (1) the power counting of KPSS breaks down in the important resonant regime and misses the generation of diffusion tails, and (2) this leads to substantial underestimation of the final matter-antimatter asymmetry by KPSS.

In summary, we have identified a new source of matter-antimatter asymmetry and set up the foundations for robust quantum transport computations relevant for a large class of electroweak baryogenesis models.

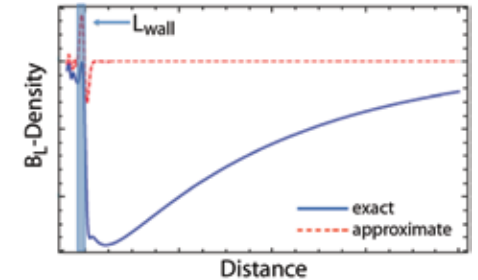


Fig. 4. Exact and approximate L-squark density profiles versus distance from the bubble wall.

[1] Sakharov, A.D., *Pisma Zh Eksp Teor Fiz* **5**, 32 (1967).

[2] Cohen, A.G. et al., *Ann Rev Nucl Part Sci* **43**, 27; arXiv:hep-ph/9302210 (1993).

[3] Cirigliano, V. et al., *Phys Rev D* **81**, 103503; arXiv:0912.3523 [hep-ph] (2010).

[4] Cirigliano, V. et al., *Phys Rev D* **84**, 056006; arXiv:1106.0747 [hep-ph] (2011).

[5] Konstandin, T. et al., *Nucl Phys B* **738**, 1; hep-ph/0505103 (2006).

# Light Element Opacities of Astrophysical Interest from ATOMIC

James P. Colgan,  
David Kilcrease,  
Norman H. Magee, Jr.,  
Gregory Armstrong,  
Joseph Abdallah, Jr.,  
Manolo Sherrill, T-1;  
Christopher J. Fontes,  
Honglin Zhang,  
Peter Hakel, XCP-5

We present new calculations of local-thermodynamic-equilibrium (LTE) light-element opacities from the LANL ATOMIC code [1] for systems of astrophysical interest. Our calculations, which include fine-structure detail, represent a systematic improvement over previous LANL opacity calculations using the LEDCOP legacy code [2]. The ATOMIC code uses ab initio atomic structure data computed from the CATS code, which is based on Cowan's atomic structure codes [3], and photoionization cross-section data computed from the LANL ionization code GIPPER [4,5]. ATOMIC also incorporates a new equation-of-state (EOS) model based on the chemical picture [6]. Our new calculations are made for elements of astrophysical interest and for a wide range of temperatures and densities.

Opacities are a crucial component in many areas of astrophysical modeling and are important to the programmatic mission of LANL. The opacities of the constituent elements of stars play an important role in the radiation transport through the stellar core and outer regions. Over the last 30 years, several worldwide groups have made great progress in generating the required opacities for elements of astrophysical interest. Collaborators within the Opacity Project (OP) have generated opacities for many elements [7,8,9]. The OPAL opacity database at LLNL [10] has provided mean opacity data for several sets of astrophysically relevant mixtures. The LANL OPLIB database has also provided opacity tables for wide ranges of temperatures and densities for many years, with the most recently available tables containing opacities generated using the LEDCOP code [2].

Recently, we have completed new opacity tables for H through Ne using the ATOMIC code and are finalizing a new table for Fe. We aim to eventually supplant the opacity data in the OPLIB database that was generated from the legacy LEDCOP calculations, the last of

which were made over 10 years ago, with data from our new ATOMIC calculations.

In Fig. 1 we compare our new hydrogen opacities with OP (short-dashed black lines) and OPAL (dashed red lines) calculations. The comparisons are made for a series of constant  $\log(R)$  values, where  $R = \rho/(T_0)^3$ , and  $T_0 = 10^{-6}T(K)$ , for  $\log(R)$  values from  $-8$  to  $+1$ . We plot the logarithm of the Rosseland mean opacity as a function of  $\log(T)$ ,

with  $T$  the temperature (in Kelvin). Several features are evident from this comparison. Overall, excellent agreement is found between all the calculations, which encompass a wide range of temperatures and densities. The new ATOMIC results generally cover a wider range of temperatures than the OP and OPAL results.

There are several regions where differences between the methods may be found. At the highest temperatures of around 8.6 keV [corresponding to a  $\log(T)$  of 8], there are some sizable differences between the ATOMIC and OPAL results. The OPAL data in this region are not particularly smooth as the temperature increases. In this fully ionized region, electron degeneracy effects and Pauli blocking are important, and the ATOMIC calculations incorporate these effects in a consistent manner. It is unclear how such non-ideality effects are included in the OPAL calculations.

At lower temperatures around 15–20 eV [ $\log(T)$  around 5.2], we also observe some differences between the ATOMIC calculations and the OP and OPAL data (which also differ somewhat from each other). These differences are evident for only the highest density cases considered. The differences are due to small differences in the EOS model used, which causes the populations of excited states of the H atom to be different. In this region, the average ionization of the plasma is just under 1 (around 0.98), and most of the opacity arises from bound-free absorption from the weakly populated states of neutral H, which are very sensitive to the EOS model used. As the temperature increases, the plasma becomes fully ionized, and the three methods are again in excellent agreement.

In Fig. 2 we turn to a comparison of a monochromatic opacity, in this case for Fe. We plot the opacity for Fe for a single temperature (15.3 eV) and for a mass density of  $5.5 \times 10^{-3} \text{ g/cm}^3$ . In this comparison we also present a LEDCOP calculation as shown by the red curve. We

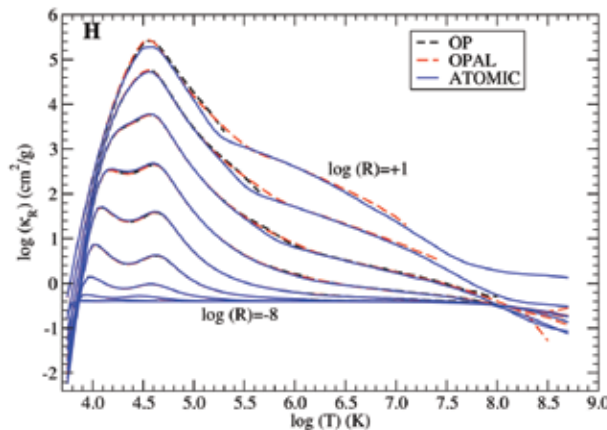


Fig. 1. Rosseland mean opacity of H as a function of  $\log$  temperature ( $T$ ) (in K). Each set of curves represents a constant  $\log(R)$  value ranging from  $-8$  to  $+1$  as labeled. We compare the current ATOMIC calculations (solid blue curves) with OP calculations (short-dashed black curves) and OPAL calculations (dashed red curves).



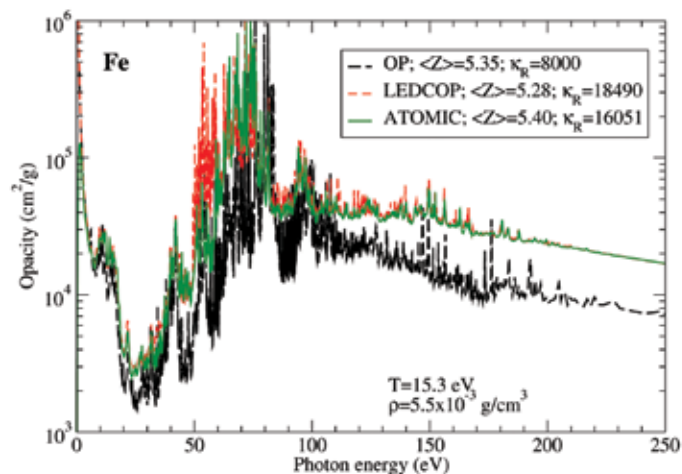


Fig. 2. Fe monochromatic opacity as a function of photon energy (in eV) for a temperature of 15.3 eV and a mass density of  $5.5 \times 10^3 \text{ g/cm}^3$ . We compare the current ATOMIC calculations (solid green curve) with an OP calculation (dashed black curve) and a LEDCOP calculation (short-dashed red curve).

find that the OP monochromatic opacity is substantially lower (around a factor of two) than both the ATOMIC and LEDCOP calculations. The Rosseland weighting function peaks at around 60 eV and in this region the OP calculation is lower than the ATOMIC

calculation by at least a factor of two. We also observe that at higher photon energies, where the opacity is dominated by bound-free contributions, the OP calculation is roughly a factor of two lower than both the ATOMIC and LEDCOP calculations. We note that these conditions are similar to those studied by other workers [11], where a variety of opacity calculations were presented. We find that the new ATOMIC calculations appear very similar to the hybrid calculations presented in [11], and similarly good agreement is found for the other temperatures and densities discussed in [11].

The ATOMIC and LEDCOP calculations presented in Fig. 2 are quite similar and produce a Rosseland mean opacity ( $\kappa_R$ ) that agrees to within 15%. The primary reason for the 15% difference is the larger opacity observed in the LEDCOP calculation at around 55 eV. The opacity in this region is dominated by  $\Delta n=0$  transitions within the 3p-3d transition array for several ion stages. Further investigations show that differences in the atomic structure calculations used in ATOMIC and in LEDCOP appear to be responsible for the larger LEDCOP opacity. The ATOMIC calculations include intermediate coupling within a configuration in the CATS structure calculations for a transition array. The LEDCOP

calculation uses algebraic splitting of the configurations into terms, a procedure that does not account for intermediate coupling. This procedure produces oscillator strengths for the lower energy transitions within the 3p-3d transition array that are substantially larger than the oscillator strengths obtained from the intermediate coupling calculation, resulting in a larger overall opacity. We regard the ATOMIC calculations as more accurate since they are based on more accurate atomic structure calculations.

In this highlight we have discussed new LANL calculations of LTE opacities using the ATOMIC code. We have completed calculations for the elements H through Ne and the calculation of our new opacity table for Fe is well underway. We plan to perform more comprehensive calculations for all elements up to Zn in the next few years.

- [1] Magee, N.H. et al., *14th Topical Conference on Atomic Processes in Plasmas*, Eds. J.S. Cohen, S. Mazevet, D.P. Kilcrease, AIP, New York, 168 (2004).
- [2] Magee, N.H. et al., *Astron Soc Pac Conf* **78**, 51 (1995).
- [3] Cowan, R.D., *The Theory of Atomic Structure and Spectra*, University of California Press, Berkeley (1981).
- [4] Clark, R.E.H. et al., *Astrophys J* **381**, 597 (1991).
- [5] Abdallah, J. Jr. et al., *J Quant Spect Radiative Tran* **71**, 107 (2001).
- [6] Hake, P. and D.P. Kilcrease, *14th Topical Conference on Atomic Processes in Plasmas*, Eds. J.S. Cohen, S. Mazevet, D.P. Kilcrease, AIP, New York, 190 (2004).
- [7] Seaton, M.J. et al., *Mon Not R Astron Soc* **266**, 805 (1994).
- [8] Seaton, M.J. and N.R. Badnell, *Mon Not R Astron Soc* **354**, 457 (2004).
- [9] Badnell, N.R. et al., *Mon Not R Astron Soc* **360**, 458 (2005).
- [10] Rogers, F.J. and C.A. Iglesias, *Astrophys J* **579**, 507; <http://opalopacity.llnl.gov/> (1992).
- [11] Gilles, D. et al., *High Energy Density Physics* **7**, 312 (2011).

# Using Neutron Star Observations to Constrain Nuclear Physics

Joseph A. Carlson,  
Stefano Gandolfi, T-2

Astrophysical observations provide important constraints to nuclear physics. In the last few years, the observation of masses and radii of neutron stars became accurate enough to be directly used to test theoretical predictions of the neutron matter equation of state (EOS). On the other side, the combination of theoretical models with recent observations opens the way to the measurement of the nuclear symmetry energy and its slope with a higher accuracy than recent terrestrial experiments.

In the recent years, important progress has been made in the observation of neutron star properties. The accuracy of these observations is giving us the confidence to provide new constraints in nuclear physics. A very important achievement is the possibility of measuring the mass and radius of neutron stars simultaneously [1] because this opens the way to extract direct information on the neutron matter EOS around nuclear densities.

Up to moderate densities of about  $2-3\rho_0$ , being  $\rho_0=0.16\text{ fm}^{-3}$ , the nuclear saturation density, neutron star matter is composed of a gas of neutrons and few protons. The easier non-trivial model to study neutron star matter is provided by the EOS of pure neutron matter. On the theoretical

side, the development of Quantum Monte Carlo (QMC) methods opened the way to study properties of nuclear systems using nuclear Hamiltonians that reproduce scattering data and properties of light nuclei with high precision [2]. Using modern supercomputers the EOS of pure neutron matter can now be calculated ab initio in a fully non-perturbative scheme [3]. Once the EOS of neutron matter is known, it is possible to study the structure of neutron stars in terms of their mass and radius by solving the Tolman-Oppenheimer-Volkoff equations.

Recently, several works have been aimed at showing the connection between the neutron

matter EOS and the symmetry energy—that is, the difference between nuclear matter and neutron matter energy. The symmetry energy is the energy cost of the isospin-asymmetry in the homogeneous nucleonic matter. Several experimental facilities aim to measure the value of the

symmetry energy, which is a great challenge. Even without knowing the experimental setup in detail, a naive idea for explaining the difficulty of these experiments is that any kind of experiment measure properties of nuclei where the number of neutrons is similar, or in a moderate excess with respect the number of protons. As a consequence, any attempt to extrapolate the results to the limit of infinite neutrons will make the result model-dependent. In the past year, we performed QMC simulations to investigate the role of the three-neutron force in neutron matter [4]. In our work we discussed the uncertainty of the three-neutron force model, and studied the uncertainty played in the EOS of neutron matter. This uncertainty pertains to the properties of neutron stars once the EOS is used as input. An important aspect of our work is that the uncertainty of the model of the three-neutron force is smaller than the current accepted range of the symmetry energy in terrestrial experiments. This is quite an important statement because the main systematic uncertainties will not be related to the model of neutron matter that we use. The EOS of neutron matter computed using QMC is shown in Fig. 1 [4]. Each band corresponds to a class of EOSs with a symmetry energy indicated in the legend and is compared to the result obtained with a two-body force only and combined with the Urbana IX three-body force.

The work in [4] suggests a convenient functional form of EOS useful to extract the symmetry energy and its slope around nuclear densities. This form of nuclear EOS has been combined with phenomenological models describing the high-density part of the EOS, and the various free parameters have been fit to the most recent neutron star observations [5]. The profile of a neutron star in terms of its mass-radius is shown in Fig. 2. The red and black bands correspond to the neutron star profile obtained from observations using two different models of the stellar matter at high-density, at the  $1\sigma$  and  $2\sigma$  (dashed lines). The observations are compared to the theoretical predictions obtained from the EOSs of Fig. 1.

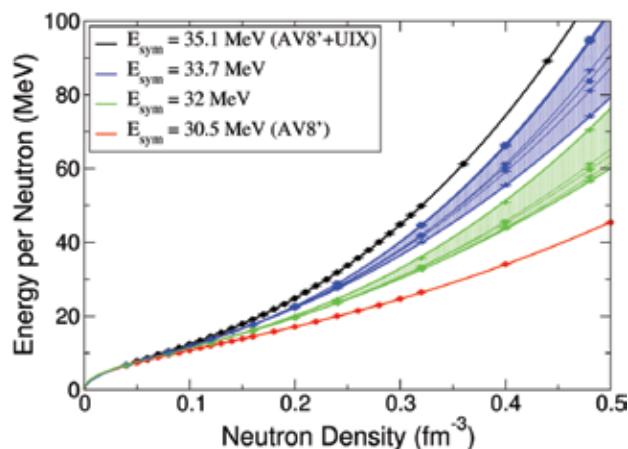


Fig. 1. The energy per particle of neutron matter corresponding to EOSs with symmetry energy indicated in the legend. Figure taken from [4].

The parameters extracted from the observations permit the derivation of the symmetry energy and its slope around the saturation density  $\rho^0$ . The results are almost model-independent, and the associated error bars are smaller than the more recent experimental data. In Fig. 3 we show a comparison of  $L$  as a function of the symmetry energy  $S_0$  (ESYM) obtained from neutron stars (indicated as “n-star”) with the most recent experimental constraints [6].

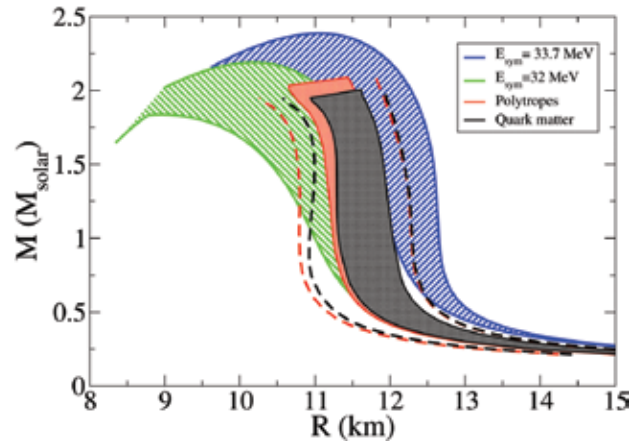


Fig. 2. The mass-radius relation of a neutron star obtained from the theoretical calculations (green and blue bands corresponding to the EOSs of Fig. 1) and from astrophysical observations red and black bands from [5].

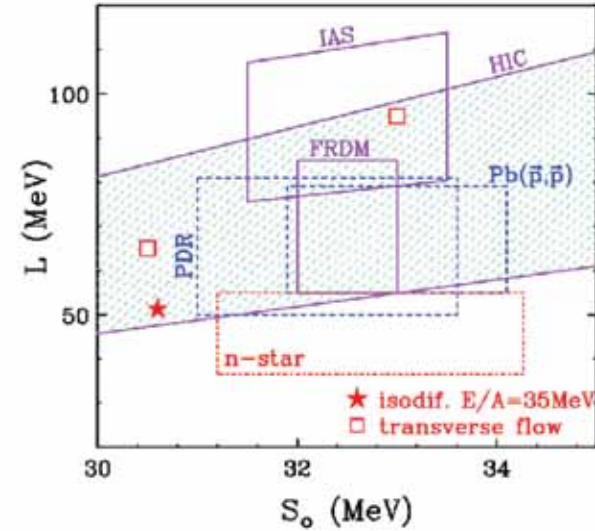


Fig. 3. Constraints of the slope  $L$  and magnitude  $S_0$  (ESYM) of the symmetry energy at saturation density from different experiments and from neutron stars observations. Figure taken from [6].

- [1] Steiner, A.W. et al., *Astrophys J* **722**, 33 (2010).
- [2] Pieper, S.C., *AIP Conf. Proc.*, **1011**, 143 (2008).
- [3] Gandolfi, S. et al., *Phys Rev C* **79**, 054005 (2009).
- [4] Gandolfi, S. et al., *Phys Rev C* **85**, 032801 (2012).
- [5] Steiner, A.W. and S. Gandolfi, *Phys Rev Lett* **108**, 081102 (2012).
- [6] Tsang, M.B. et al., *Phys Rev C* **86**, 015803 (2012).



# Direct Numerical Tests of Kinetic Theory Collision Integrals with Molecular Dynamics Simulations of Stopping Power in Plasmas

Paul E. Grabowski,  
Michael S. Murillo, CCS-2;  
Michael P. Surh,  
David F. Richards,  
A. Bruce Langdon,  
Frank R. Graziani, Lawrence  
Livermore National  
Laboratory

Molecular dynamics (MD) is used as a validation tool for different stopping models based on kinetic theory. MD is also used to improve those models so they are accurate over a wider range of conditions. Unlike many transport quantities, stopping is difficult to accurately model because one must reproduce a function of projectile velocity instead of an average over a velocity distribution. Our new models are accurate to within a few percent. These models can now be used to derive general transport coefficients.

All of transport theory (diffusion, thermal conductivity, electrical conductivity, stopping power, etc.) of non-zero mass particles is derived from kinetic theory, in which a phase space distribution function is evolved in time. Calculating the correct evolution is dependent on having the right collision operator that models how small-distance and small-time interactions between particles change, and eventually thermalize, the distribution. MD simulations provide an ideal numerical laboratory in which different classical kinetic theories can be tested and broken by directly simulating all inter-particle interactions, and hence the full many-body phase-space distribution.

The mean energy loss of a projectile per unit distance ( $dE/dx$ ), or stopping power, in a classical target was chosen as our observable of interest for three reasons: (1) it is the only velocity-resolved transport property, (2) the classical linear response theory leads to a divergent answer without an ad hoc cutoff at small distances, and (3) stopping power is itself an

important property relevant to thermonuclear fusion in inertial confinement fusion devices, for both alpha particles produced during burn and for fast ignition heating of the target. Therefore, stopping power is both the most difficult transport quantity to model correctly and of major relevance to the core mission of LANL.

MD simulations were performed with the two-time Gordon Bell prize winning code ddcMD, originally developed at LLNL and now ported to the LANL supercomputer, Cielo. The code ddcMD is capable of simulating billions of particles with long-range forces and is a virtual plasma laboratory and workhorse code of the Cimarron project [1]. Recently, it has been used to study electron-ion temperature equilibration [2], fusion enhancement rates [3], and ignition [4] while thermal conductivity and interface diffusion results are in progress.

We have studied a wide range of conditions, varying the Coulomb coupling parameter of the target,  $\Gamma_t = e^2/ak_B T$ , a measure of the ratio of mean potential energy  $e^2/a$ , where  $a$  is the electron sphere radius, to mean kinetic energy  $k_B T$  in the system, from  $\Gamma_t = 0.1$  (near pure hydrogen plasma at thermonuclear burn conditions) to  $10.0$  (stopping power experimental conditions), and using three different types of projectiles: (1) anti-protons ( $Z = -1$ ), (2) anti-alpha particles ( $Z = -2$ ), and (3) anti-Neon nuclei ( $Z = -10$ ). Note that because the projectiles have the same charge as the electrons, we do not study the bound component of stopping.

Figure 1 shows the electron density,  $\rho$ , around a moving anti-alpha particle ( $v \sim 9 v_{th}$ ), and a color image of the energy deposition field,  $\varphi$ , is superimposed. This is the differential work performed on the target electrons by the projectile. Both fields are time and azimuthal averages of the instantaneous particle distributions. For this case, the major part of the energy transfer occurs in front of the projectile via strong particle-particle collisions with the electrons in the target. The dynamic collective response leads to damped plasma oscillations in the tail behind the moving projectile. These are visible in the grayscale density  $\rho$  and in the faint alternations in the energy flow from projectile to target. The wavelength in the wake scales as the product of the projectile velocity and plasma oscillation. In the future, we will generalize this to a two-component plasma in which separate current-density fields of electrons and fuel ions will be used to compute the electron-fuel energy split.

Figure 2 shows the unitless low velocity friction coefficient  $R = \lim_{v \rightarrow 0} (dE/dx) \lambda^2 v_{th} / v Z^2 e^2$ , where  $\lambda$  is the Debye length,  $v_{th}$  is the thermal velocity,  $v$  is the velocity of the projectile and  $Ze$  is the projectile charge. This quantity is related to the diffusion coefficient [5], which can be obtained via a fit to molecular dynamics [6]. Another model valid at low velocities is the T-matrix model, which depends on a cross-section through the Boltzmann collision integral. The use of the Coulomb

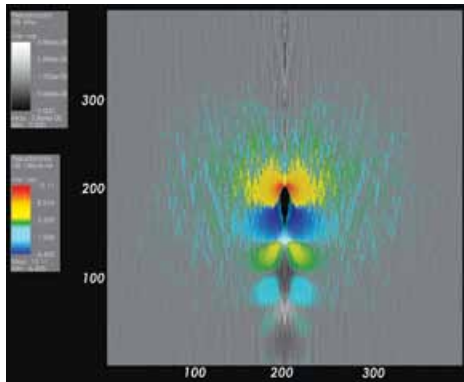


Fig. 1. The wake field is shown for a  $Z=-2$  projectile near the center of the image, moving in the positive  $y$  direction. The field of view is  $720 \text{ \AA}$ , compared to the cell length of  $1075 \text{ \AA}$ ; the screening length is  $2.4 \text{ \AA}$ . The electron charge density,  $\rho$ , is shown in gray; the strongly-repulsive screening cloud is the dark lozenge. The logarithm of the energy transfer field,  $\varphi$ , is shown in color. (Specifically,  $\text{Sign}(\varphi) \ln |\varphi|$  is shown with a cutoff at small values.) The result is time-averaged over the 400 fs duration of the simulation and is cylindrically-averaged around the axis of the particle trajectory.

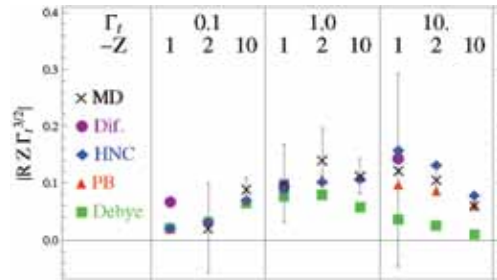


Fig. 2. The friction coefficient over a range of different Coulomb coupling parameters and projectile charges. Results are shown for MD (black crosses with error bars), the diffusion model of Dufty (purple circles), and T-matrix with a screening potentials derived from hypernetted chain (blue diamonds), the Poisson-Boltzmann equation (red triangles), and Debye-Huckel theory (green squares).

cross-section would be unphysical because real interactions are screened by the medium; furthermore, that result would be divergent. The screening cloud about the projectile can be calculated to high accuracy with the hypernetted chain approximation (HNC) [7]. Alternatively, one can solve the non-linear Poisson-Boltzmann (PB) equation, which assumes a Boltzmann

distribution of electron energies and Coulomb interactions between particles. The linear version of this equation leads to simple Debye screening. Cross-sections from all three of these screening models are used and compared. We see that at small coupling, the choice of potential does not matter, but at large coupling the Debye potential will significantly underestimate the stopping.

The full stopping curve is a function of the velocity of the projectile. Static models break down when  $v$  is about  $v_{th}$  or greater, so a dynamic screening model is needed. Within linear response, one can write the random phase approximation (RPA) to the dielectric function and calculate the stopping, but this model breaks down at low velocities, where binary collisions make up the majority of the stopping. Gould and DeWitt [8] corrected for this behavior by adding in the T-matrix result and then subtracting the static RPA result in order to avoid double counting. This model works well at small coupling but breaks down at strong coupling. Zwicknagel [9] instead used the T-matrix method with the slight change of scaling the screening length with velocity to match the high-velocity Bohr limit. This works very well, but tends to underestimate the stopping by 10–15% (and more at low velocities) for moderate to strong coupling. We have made two improvements to this model: (1) we use more accurate potentials in the T-matrix model, and (2) we have altered the scaling factor to better match the MD data at moderate to high velocities. Figure 3 compares these models at small and large coupling.

In summary, MD is a powerful tool for validating models that are used in hydrodynamics and kinetic theory codes. We have used it here to test stopping models over a wide range of conditions and to develop improved models able to match the MD data to within a couple of percent. These models are now validated at the velocity resolved level and can be used more generally in other transport calculations.

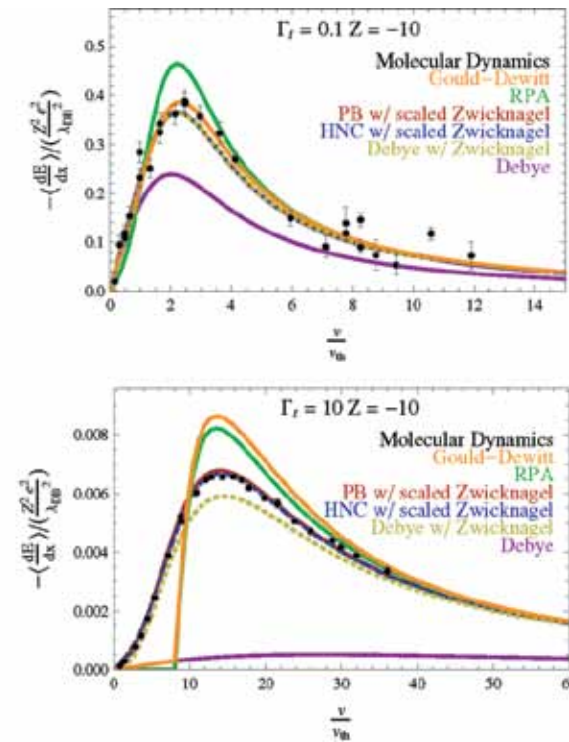


Fig. 3. The stopping power as a function of velocity for an anti-Neon projectile in a weakly and strongly coupled electron target. Results are shown for MD (black dots with error bars), the combined model of Gould and DeWitt (orange), the random phase approximation (green), and T-matrix with a scaled screening potential derived from the PB equation (red), hypernetted chain (blue), and Debye-Huckel theory (yellow), and the unscaled Debye-Huckel potential (purple).

- [1] Graziani, F.R. et al., *High Energy Density Physics* **8**, 105 (2012).
- [2] Benedict, L.X. et al., *Phys Rev E* **86**, 046406 (2012).
- [3] Fichtl, C.A. et al., *TSC Science Highlights*, **96** LA-UR-12-20429 (2012).
- [4] Glosli, J.N., *Computational Challenges in Hot Dense Plasmas Workshop*, Institute for Pure and Applied Mathematics, www.ipam.ucla.edu/abstract.aspx?tid=10446 (2012).
- [5] Dufty, J.W., *Nucl Instrum Meth Phys Res B* **96**, 626 (1995).
- [6] Ohta, H. and S. Hamaguchi, *Phys Plasma* **7**, 4506 (2000).
- [7] Balescu, R., *Equilibrium and Non-equilibrium Statistical Mechanics* (1975).
- [8] Gould, H.A. and H.E. DeWitt, *Phys Rev* **155**, 68 (1967).
- [9] Zwicknagel, G. et al., *Phys Rep* **309**, 117 (1999).

# The New Finite-Range Droplet-Model Mass Table FRDM(2012) and Associated Tables of $\beta$ -Decay Properties

Peter Möller, T-2

About 100 elements exist in nature. Each element is defined by its proton number, but can exist as different isotopes depending on neutron number. Only a few different, stable isotopes of each element exist, but about 100 different isotopes of many elements can exist in environments such as stars. To understand and model processes in stars and in nuclear reactors we need to know nuclear properties such as masses and decay half-lives. Because of the instability of many of these isotopes, their properties have not been experimentally measured; instead we model them. Our previous “edition” of calculated masses and decay half-lives, giving these properties for about 9,000 nuclides, is from 1992 and has been extensively used (cited about 2,000 times) in many simulation areas. By use of improved computers and enhanced physics modeling we have now developed a new and better database of nuclear masses and decay half-lives, FRDM(2012).

Our previous mass table edition, finalized in 1992 and published in 1995 [1] reproduced 1,654 experimental binding energies (masses) to an average accuracy of 0.67 MeV. Binding energies for heavy nuclei are about 1,500 MeV so this represents an accuracy to within 0.04%. The table was calculated on Cray computers. To perform the identical calculation today would take only about three hours on a modern laptop. Our current edition FRDM(2012) has taken about 50,000 CPU hours on a modern cluster. This is fairly modest by some standards, but the reasonable demand on computer time has permitted us to investigate many plausible ideas and, after thorough studies, select the optimal approaches. The steps that have taken us from FRDM(1992) to FRDM(2012) are summarized in Fig. 1. The first three steps are discussed in [2], the inclusion of the axial asymmetry shape-degree of freedom, step 4, in [3], and step 5, the effect of considering density-symmetry effects, in [4]. Step 6, improved ground-state correlations, will be discussed in a future publication. After the six successive steps were implemented, we arrived at the new mass model (table) FRDM(2012), which includes 9,318 calculated masses and reproduces 2,149 known masses in a 2003 experimental database [5] to an accuracy of better than 0.56 MeV.

In Fig.2 we compare the previous and current mass tables to experimental data. To an eye unfamiliar with the subject area, the improvement may not seem substantial. However, we note that the sharp cusp in the displayed lines around  $N=126$  are gone, the large fluctuations in the vicinity of  $N=82$  are significantly diminished, and the (negative) bump just above  $N=40$  is removed. Also, the lines are in general closer together and closer to zero, in particular above  $N=60$ . All

these features can be expected to improve modeling of nucleosynthesis in the r-process. We have indeed found that this is the case in a first, preliminary study based on the new mass database and the associated new calculated database of  $\beta$ -decay properties. Initial results show that the new databases yield important improvements in the agreement between calculated r-process abundances and the observed abundances in the sun, and improve our understanding of the r-process in general [6]. It is particularly valuable that we have calculated nuclear masses and  $\beta$ -decay properties within a single, consistent microscopic model. This facilitates enormously the interpretation of what are the important physics processes governing the phenomena we study.

The r-process studies provide somewhat indirect evidence that we model well the properties of nuclei where the r-process occurs, which is far from stability and far from known nuclei. These nuclei have many more neutrons than nuclei that can currently be studied experimentally, in some cases 20 or so more neutrons. We can also ask what is more direct evidence that results are improved? We show in Fig. 3 the differences between the old and new calculations versus proton and neutron numbers. The plot covers about 9,000 nuclides. Most are actually unknown, but we can calculate their properties in our model. For large neutron numbers there is a growing, systematic difference. Since it is in an unknown region of nuclei, at the end of the r-process we cannot say how realistic the new results are. More testable are the “islands” of fairly large differences between the new and previous mass tables, when they occur where experimental masses are available. We have been able to verify that the new results are in better agreement with experiment. Let us focus on the region near proton number 40 and neutron number 40. In this region there are new experiments with newly measured masses

Successive FRDM enhancements	
<b>Optimization (2006)</b>	
Better search for optimum FRDM parameters.	
Accuracy improvement:	0.01 MeV
<b>New mass data base (AME2003) (2006)</b>	
Better agreement than with AME1989.	
Accuracy improvement:	0.04 MeV
<b>Full 4D energy minimization (2006–2008)</b>	
Full 4D minimization( $\epsilon_2, \epsilon_3, \epsilon_4, \epsilon_6$ ) step=0.01.	
Accuracy improvement:	0.02 MeV
<b>Axial asymmetry (2002–2006)</b>	
Also yields correct SHE gs assignments.	
Accuracy improvement:	0.01 MeV
<b><math>I, \pi</math> variation (2009–2011)</b>	
Accuracy improvement:	0.02 MeV
<b>Improved gs correlation energies (2012)</b>	
Accuracy improvement:	0.01 MeV

Fig. 1. Impact of successive enhancements to FRDM(1992) with  $\sigma = 0.669$  MeV, leading to FRDM(2012) with  $\sigma = 0.5594$  MeV. The years in parentheses indicate when this particular study took place.



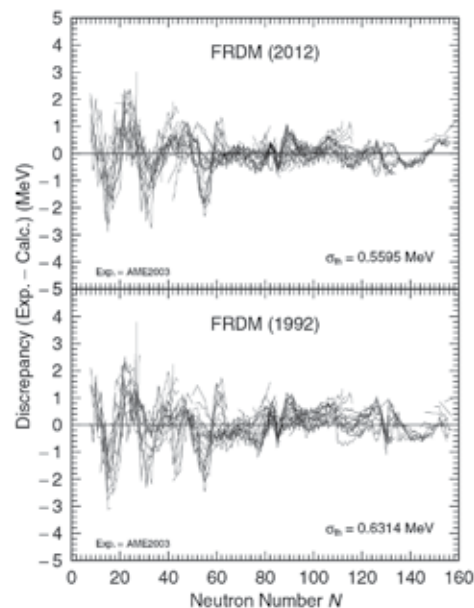


Fig. 2. Difference between experimental and calculated masses for the previous and current model. Lines connect isotopes of the same elements.

available to which the model parameters were not adjusted. In Fig.4 we compare the previous mass table (FRDM(1992)) and the current one (FRDM(2012)) to the complete set of newly measured masses [7]. Whereas the previous mass table showed large deviations in this region, with an average deviation between calculated and experimental masses of 1.10 MeV, in the new mass model the accuracy is improved to 0.34 MeV. It is outside the scope of our presentation here to discuss precisely why the new model improved the masses in this and other localized regions; suffice it to say that improved computer power permitted us to abandon some approximations made earlier and explore the model more satisfactorily; the precise reasons for the improvements are fully understood.

Based on our initial benchmarking of the new mass table we expect it will contribute to improved modeling of nucleosynthesis in stars and of many other phenomena.

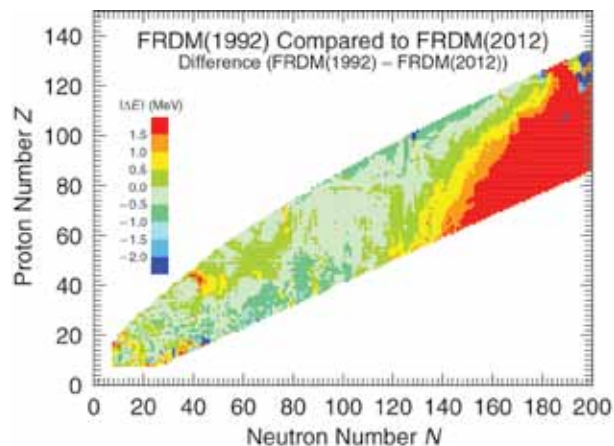


Fig. 3. Difference between the previous and current mass model masses.

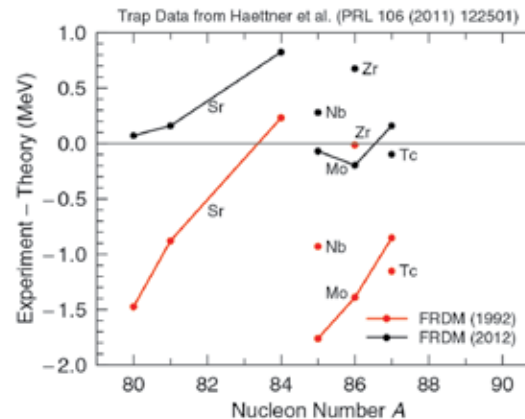


Fig 4. Difference between a complete set of masses measured in a specific recent experiment and the predictions of (1) the previous mass model FRDM(1992) and (2) the new mass model FRDM(2012). The accuracy improved from 1.10 MeV to 0.34 MeV. The newly measured masses were not used in the determination of model parameters.

[1] Moller, P. et al., *Atomic Data Nucl Data Tables* **59**, 185 (1995).

[2] Moller, P. et al., *Proc Int Conf Nuclear Data Technol*, **69**; ISBN978-2-7598-00902; <http://t2.lanl.gov/molleretal/publications/nd2007.html> (2008).

[3] Moller, P. et al., *Phys Rev Lett* **97**, 162502 (2006).

[4] Moller, P. et al., *Phys Rev Lett* **108**, 052501 (2012).

[5] Audi, G. et al., *Nucl Phys A* **729**, 337 (2003).

[6] Farougi, K. and Kratz, K.L. Priv. Communication.

[7] Haettner, E. et al., *Phys Rev Lett* **106**, 122501 (2011).



## Chemistry and Biology

LANL has a long history of outstanding work in chemistry and biology. This work is dependent upon the collaborations between theorists and modelers in ADTSC and experimentalists here at LANL, as well as colleagues in academia and other national laboratories. The research in these areas often has national security applicability while advancing basic science.

Articles in this section cover research in protein-DNA binding, epidemiology, genetic material analysis, early indications of infectious disease, material interfaces, RNA-based molecular switches, and electrochemical stability of ionic liquids.



# New Insights into Protein-DNA Binding from Simulations of DNA Breathing Dynamics

**Boian S. Alexandrov, T-1;**  
**Amy L. Bauer, XTD-4;**  
**Alan R. Bishop, PADSTE; William S. Hlavacek,**  
**Fangping Mu, T-6**  
**Elizabeth Hong-Geller, B-10;**  
**Kristy Nowak-Lovato, D-3;**  
**Kim Ø. Rasmussen, T-5;**  
**Ludmil B. Alexandrov, Wellcome Trust Sanger Institute**  
**Anny Usheva, Harvard Medical School**

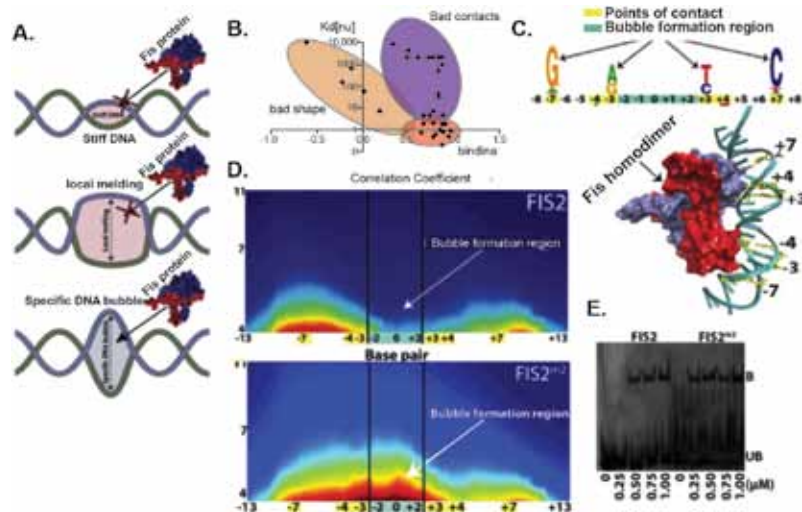
Understanding the mechanisms governing the affinities of transcription factors (TF) for their DNA binding sites is of tremendous importance for the regulation of gene expression in the cell. The DNA physicochemical properties that are most relevant to the TF-DNA binding have yet to be fully understood. Recently, we investigated the role of DNA breathing dynamics as a mechanical property of DNA regulating TF-DNA binding, and specifically the relationship between the DNA local propensity for breathing and binding of two well-known TFs: (1) the human TF YY1, and (2) the nucleoid-associated protein Fis in *Escherichia coli*. Our results demonstrate that suppressing or enhancing the DNA propensity for local transient opening at known binding sites, for example, by introducing a few base pairs (bp) substitutions or nucleotide modifications such as methylation (designed by computer simulations to change DNA breathing without changing the direct points-of-contact) can drastically affect TF binding.

**Fig. 1. A:** Schematic representation of DNA bubble formation in Fis-DNA binding—(i) Weak Fis-DNA binding to stiff sequences; (ii) Weak Fis-DNA binding to significant destabilization of the double helix; and (iii) Strong Fis-DNA binding in the presence of a characteristic breathing profile. **B:** Schematic affinity shape-correlation diagram of *in vitro* sequences. Each point represents an oligomer characterized by 1) specific direct points-of-contact, 2) correlation with the characteristic opening profile, and 3) dissociation constant. **C:** The Fis logo and crystal structure example of Fis-DNA binding complex [8]. **D:** Langevin dynamics simulations demonstrating enhanced FIS2m2 local DNA breathing dynamics. FIS2m2 was designed by introducing two O6-methylguanine in the bubble formation region of the FIS2 sequence, without changing the points-of-contacts [7]. The color map represents the probability of bubble opening; high (red) and low (blue) probability of opening. The length of the transient bubbles, in base pairs [bp], is shown along the vertical axis. The horizontal axis depicts bp position; the bubble formation region is in blue while the points-of-contact are in yellow. **E:** EMSA demonstrating the increase in complex formation of FIS2m2 versus FIS2. Adapted from [7].

Cellular proteins are TFs that regulate gene expression, and thereby cellular activities and fate, by binding to specific DNA segments positioned at different regulatory locations in the genome. Locations and specificities of a TF's binding site are determined by a complex interplay of specific direct points-of-contact (direct recognition) and by a set of physicochemical properties of DNA, such as local shape (indirect recognition). Despite the important role of TFs in myriad cellular processes and diseases, the physical mechanisms underpinning the strength of binding (affinity) and recognition (specificity) are still far from being completely understood. Recently developed high-throughput techniques, such as chromatin immunoprecipitation coupled with microarrays, next generation sequencing, or exonucleases and protein binding microarrays (PBM), have enabled large-scale profiling and characterization of binding sites and revealed a complex landscape

of DNA binding preferences [1,2]. Only a fraction of the observed TF binding sites can be explained by direct (or protein-partner assisted) binding or captured with a single consensus sequence or position weight matrix. Even the characterization of cell-specific chromatin structure and leveraging of sophisticated motif recognition protocols still left surprising exceptions in which the binding, or the lack of it, is hard to explain [2]. A possible reason is that the bioinformatic models utilized to explain and predict TF binding sites, which rely on letter-code representations of DNA sequences, neglect physicochemical features of DNA that are important for protein-DNA recognition [3]. Here we highlight two recent studies showing that a specific physical property of DNA, namely the propensity for transient opening and re-closing of the double helix, resulting from inherent thermal fluctuations, also known as “DNA breathing” or “DNA bubbles,” is associated with binding affinity, at least in the case of Fis and YY1, both of which are well-studied TFs. Using a DNA mesoscopic nonlinear model [4] that can be used for rational design of local DNA local breathing [5], we simulated breathing dynamics of known or engineered Fis and YY1 binding sites [6,7] and found a strong association between the propensity for breathing (at the binding sites) and YY1 and Fis binding affinity (at the same sites) (Fig 1, A).

**Fis:** Our simulation data and available *in vitro* Fis binding data indicate a strong correlation between DNA breathing dynamics and Fis binding (Fig. 1, B). The Fis direct points-of-contact are predominantly on one side of the double-stranded DNA binding segment [8] (Fig 1, C). It is natural to expect that the binding of a protein with direct points-of-contact predominantly on one of the two DNA strands, such as Fis, can be promoted by transient openings of the DNA at the TF binding site. Leveraging simulations of the breathing of known Fis *in vitro* binding sites, we defined an average DNA breathing profile that is characteristic of a strong



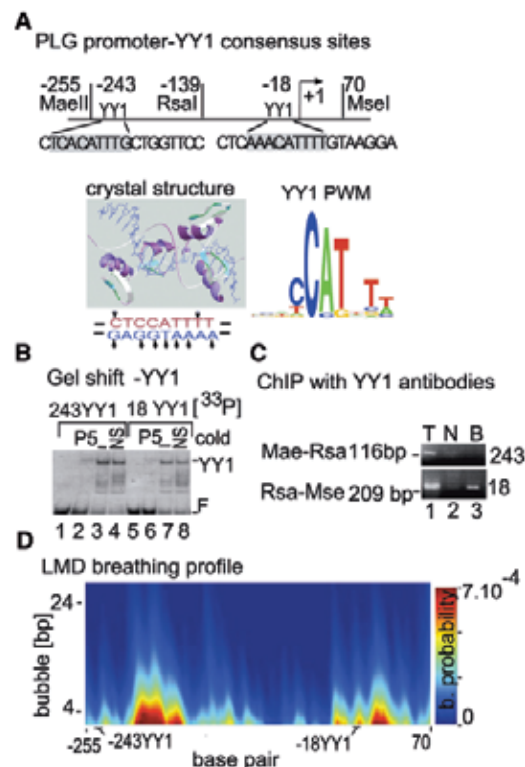


Fig. 2. A: YY1 binding sites locations relative to the transcription start site (TSS) (+1) of the human PLG promoter. The crystal structure of the YY1-P5 DNA complex [9] and the base-specific YY1 contacts on both strands (arrows). The YY1 logo is shown on the right. B: Gel shift assays demonstrate recombinant YY1 binding to 24 bp long synthetic oligonucleotides containing the two sequences: 243YY1, respectively, 18YY1 (0.1 nM), as indicated above the plots. C: ssChIP assay is used to verify genomic YY1 binding at the identified consensus sequences. After sonication and restriction enzymes digestion the YY1 antibody-captured promoter fragments are amplified by PCR with fragment-specific primers; line 1—total DNA before antibody selection (T); line 2—pulled-down DNA with control antibody (N); line 3—pulled-down with YY1 antibody (B). The identity of the PCR-amplified fragments is shown at the left: Hae-Rsa contains the 243YY1; the 18YY1 site is located in the Rsa-Mse restriction fragment. D: LMDs simulations demonstrating local DNA breathing dynamics in a 305-bp-long human PLG promoter fragment. The notations are as in Fig. 1. Adapted from [5].

Fis binding site and demonstrated that this profile is significantly enriched among known *in vivo* *E. coli* Fis binding sites. Furthermore, to test our understanding of how Fis binding is influenced by DNA breathing dynamics, we numerically designed base-pair substitutions and O6-guanine-methylation modifications in DNA segments (Fig. 1, D) that are known to interact strongly or weakly with Fis. In each case we sought to make local DNA breathing dynamics either closer to or farther from the breathing profile characteristic of a strong Fis binding site without changing the direct points-of-contact. For the modified DNA segments, we found that Fis-DNA binding, as assessed by electrophoretic mobility shift assay, is changed in accordance with our expectations (Fig. 1, E). We conclude that DNA breathing dynamics, which depends on DNA mechanics, and can be modified by nucleotide modifications such as nucleotide methylation, regulates Fis binding.

The regulation of DNA mechanics by nucleotide (i.e., cytosine, guanine, or adenine) methylation that can change the local breathing propensity of a regulatory genomic segment could play an important role in a novel gene-silencing mechanism. Such a mechanism will be based not on hypo- or hyper-methylation of the promoter segment, which leads to a dense chromatin structure [10], but on the presence of a few methylated nucleotides located at specific positions that can suppress or enhance binding of various TFs via alteration of local DNA breathing.

**YY1:** By using site-specific chromatin immunoprecipitations (ssChIP), EMSA, BIOBASE data, and our computational framework [8], we found a specific DNA breathing profile associate with YY1 binding sites *in vivo*. In particular, we found that genomic flanking sequence variations and single nucleotide polymorphisms (Fig 2, A) may exert long-range effects on DNA dynamics and thereby influence YY1 binding *in vivo*. The correlation between our simulations and ssChIP data clearly emphasize the role of the local propensity for breathing at the YY1 binding site. For example, we found that, although YY1, *in vitro*, strongly binds two short DNA segments (identical with two DNA segments at the human PGL promoter) that contain the YY1 consensus motif (Fig 2, B), the same sequences tested for YY1 binding *in vivo*, via ssChIP, demonstrated a lack of binding for one of them (Fig 2, C). We interpreted this phenomenon as an influence of the genomic flanks on the breathing of the YY1 binding sites, (Fig. 2 D). YY1 is a Zn-finger protein that binds to the major groove without conformational changes in the DNA or the protein itself [9]. The four YY1 fingers make multiple contacts with the major groove edges of bases but most of the specific hydrogen-based contacts are restricted to one of the two DNA strands (Fig. 2, A), as in the case of Fis. Therefore, it seems natural that the requirements for enhanced transient DNA openings at the points-of-contact region predicted by simulations which account for the effects of flanking sequences, would facilitate specific YY1 binding.

- [1] Badis, G. et al., *Science* **324**, 1720 (2009).
- [2] Arvey, A. et al., *Genome Res* **22**, 1723 (2012).
- [3] Maienschein-Cline, M. et al., *Nucleic Acids Res* doi: 10.1093/nar/gks771 (2012).
- [4] Bishop, A.R. et al., *Entropy-Driven Conformations Controlling DNA Functions*, Verlag-Berlin-Heidelberg, Springer (2012).
- [5] Alexandrov, B.S. et al., *Nucleic Acids Res* **38**(6), 1790 (2010).
- [6] Alexandrov, B.S. et al., *Nucleic Acids Res* **40**(20), 10116 (2012).
- [7] Nowak-Lovato, K. et al., "Binding of Nucleoid-associated Protein Fis to DNA is Regulated by DNA Breathing Dynamics," *PLOS Comp Biol*, <http://www.ploscompbiol.org/article/info:doi/10.1371/journal.pcbi.1002881> (2012).
- [8] Stella, S. et al., *Genes Dev* **24**, 814 (2010).
- [9] Houbaviy, H.B. et al., *Proc Natl Acad Sci USA* **93**, 13577 (1996).
- [10] Tate, P.H. and A.P. Bird, *Curr Opin Genet Dev* **3**(2), 226 (1993).

# Epidemiological Analysis for Outbreak Control

Nick Hengartner, CCS-3  
Ruy M. Ribeiro,  
Paul W. Fenimore, T-6

Throughout history infectious disease has been a major cause of death, disability, and suffering. Disease control in recent times is based on a rational understanding of the patterns, causes, and effects of infectious disease and has been essential to the reduction of death and disability rates to less than 10% of historical rates in the last century in the US [1]. Striking examples of the power of rational techniques are the effective extinction of the organisms responsible for smallpox and rinderpest. The obvious question is the reach of these powerful techniques for solving other similar infectious disease problems. The emergence of the H1N1 pandemic influenza in 1918 is still not universally understood as a prime example of the damage that an emerging disease can cause. This raises the question whether there are rational methods that can help mitigate or even extinguish an emerging disease before it reaches the scope of smallpox, rinderpest, the 1918 pandemic influenza, or HIV disease. Analytical methods do exist—as those that allowed prediction of the control of severe acute respiratory syndrome (SARS) and estimated the level of contagion from the H5N1 bird-flu and 2009 pandemic influenza demonstrate. These promising developments show a path toward the use of rational interventions to improve the control of novel disease outbreaks.

The rational study of patterns, causes and effects of infectious disease is known as epidemiology. For contagious disease—a subset of infectious disease—that study often focuses on understanding the dynamics of how the disease propagates within a defined population, determining how fast it spreads, estimating the total number of infected individuals, and predicting the impact of prophylactic measures. Field and theoretical epidemiology played essential roles in the elimination of smallpox and rinderpest, providing the basis for the allocation of the human effort and resources that made global eradication a tractable problem. Theoretical epidemiology showed great promise during the global 2003 SARS outbreak as a tool for determining when infection control and mitigation had slowed disease transmission sufficiently so that control of the outbreak could be achieved and for anticipating the final regional size of the outbreak.

Control of the SARS epidemic in 2003 was a signal event. The disease exhibited high lethality, showed every indication of becoming a global pandemic, and had spread sufficiently once control efforts commenced that there were serious doubts that control could be achieved and a pandemic averted. The vast majority of the effort required to control the SARS outbreak consisted of field epidemiology, contact tracing, barrier nursing, and a host of other established medical and public health techniques. These techniques averted a pandemic, but they did not provide a strong analytical basis for identifying whether and at what point in time control of the outbreak had been achieved, nor what the time frame and final case count were likely to be. A team at

LANL [2], one of several in the world, was able, with adequate access to epidemiological case-count data, to demonstrate the point in time at which eventual control of the epidemic was likely for specific countries or regions. Because the public health infrastructure and methods of control differed so strongly between countries, global predictions in the context of a single modeling exercise were not feasible in 2003.

The remarkable success of the global public health effort in controlling SARS notwithstanding, serious limitations in the deterministic methods used to analyze case-count data were brought into sharp focus in 2003. The methods used during the SARS outbreak to predict the course of the epidemic were derived from relatively simple-to-analyze ordinary differential equations, related to the minimal mathematical model,

$$\begin{aligned}\frac{dS(t)}{dt} &= -\beta \frac{S(t)I(t)}{N}, \\ \frac{dI(t)}{dt} &= +\beta \frac{S(t)I(t)}{N} - \gamma I(t), \\ \frac{dR(t)}{dt} &= +\gamma I(t)\end{aligned}$$

where  $S(t)$  is the population of susceptible individuals at time  $t$ ,  $I$  and  $R$  describe the population of infectious and recovered (or “removed”) individuals,  $\beta$  describes the chance of an infectious individual infecting a susceptible person, and  $\gamma$  is the rate of recovery (or “removal”) of infectious individuals. Generalizations of these equations were used in 2003 to predict 396 probable and suspected cases of SARS in Ontario,



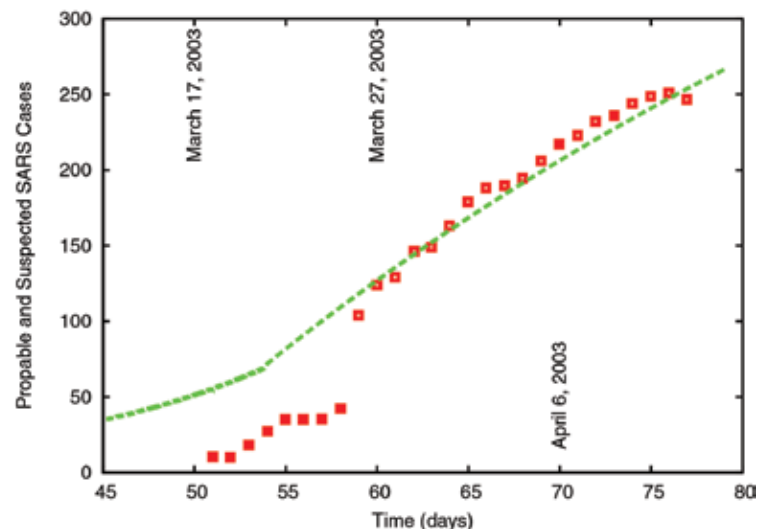


Fig. 1. The cumulative number of probable and suspected Coronavirus SARS cases during the 2003 Ontario, Canada SARS outbreak. The jump in the data near day 60 reflects real-world reporting as knowledge of the situation improved. The green curve reflects a deterministic model accounting for the effects of variable diagnosis and patient isolation effects that change the transmission of the disease and alter epidemic progression. The kink in the green curve corresponds to the start of epidemic control measures by Canadian authorities. The model used to generate the green curve is a more complicated set of equations analogous to the simple susceptible, infectious, recovered model discussed in the text. The green curve extrapolates to a final case count of 396 cases. The Canadian government reports 375 cases.

while the actual result was 375 cases (247 probable cases and 128 suspected cases). Figure 1 shows the predictions made in 2003 using real-world non-ideal data [2]. While equations of this type provide a reasonable first approximation to the epidemic dynamics, they provide no statistical confidence or likelihood for their predictions.

The deterministic model of disease spread was adequate for making important predictions in the 2003 SARS outbreak, but the uncertainties in those predictions can only be estimated in an ad hoc way, and the model is the least reliable early in the outbreak when case counts are small and prospects for controlling the outbreak are best. Ideally, we would have a statistical measure of how likely an outbreak is to grow as an epidemic and what the likely growth rate is to help estimate the efficacy of disease control measures in real time. The human H5N1 “bird flu” outbreaks that have simmered for over a decade have resulted in several dozen confirmed deaths each year for most of the last decade. The reason these outbreaks result in locally confined outbreaks—not global pandemics—is that the virus has failed to achieve efficient person-to-person transmission. Using Bayesian statistical methods to

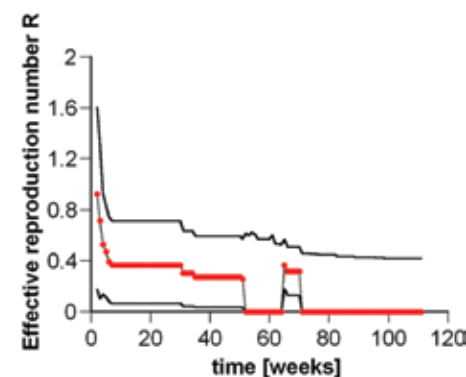


Fig. 2. Sequential Bayesian estimation of the basic reproductive number  $R$  for H5N1 bird-influenza infecting people in Vietnam. The mean estimated value of  $R$  is shown in red, and the 95% credible interval lies between the upper and lower black curves. Values of  $R$  smaller than 1 indicate an outbreak is no longer viable and that spread will be self-limiting.

generalize the deterministic equations to a probabilistic form, we can estimate the number of new cases caused by each current case. This allows us to estimate the range of potential epidemic growth rates most consistent with the observed case counts (Fig. 2) [3]. This technique is applicable even very early in an epidemic when case counts are small. This represents a significant advance in our ability to provide high-quality information on the effectiveness of disease control efforts to public health officials during a crisis. Knowing the range of potential growth rates means that a more realistic view of the possibility for epidemic growth can be achieved than the simple picture of “control/no control” that was available during the SARS outbreak.

[1] Centers for Disease Control, *Morbidity and Mortality Weekly Report* **48**, 621 (1999).

[2] Chowell, G. et al., *J Theor Biol* **224**, 1 (2003).

[3] Bettencourt, L.M.A. and R.M. Ribeiro, *PLoS One* **3**, e2185 (2008).

# Making the Dark Matter of Biology Visible

Joel Berendzen, P-21;  
Mira Bussod, T-6;  
Judith Cohn,  
Nick Hengartner, CCS-3;  
Ben McMahon, T-6

The advent of large-scale shotgun sequencing of entire microbial communities, viewed as a meta-organism, promises to make the dark matter of biology visible by helping us understand microbial communities important to medicine, ecology, industry, or agriculture [1]. The basic problem in using shotgun sequencing data from environmental samples is to roughly identify the phylogenetic origin and possibly the gene each sequence read, often only 100–200 bp long, belongs too. The current approach is to compare each read to the entire library of sequenced data. While Genbank contains a truly enormous quantity of genetic sequence data to which comparisons can be made (over 1,500 distinct bacterial species, along with approximately 100 fungi, 100 archaea, 50 metazoa, and 50 plants and protozoa), its coverage of organisms on the tree of life is uneven. A cursory identification of genomic fragments can be made by assessing its similarity to elements in the database, using versions of the blast algorithms developed at LANL in the 1980s. This comparison is computationally cumbersome because it may take several weeks on large computer clusters. This computational burden is the main bottleneck for extracting information from the shotgun sequencing of environmental samples.

A game-changing approach that enables us to analyze the genomic content from environmental samples is to look for meaning instead of similarities in the sequence reads. The elements required for success in analyzing languages are also present in the genetic language of DNA. A typical bacterial genome encodes for several thousand proteins, with a total of around one million amino acids involved in their sequences. Like the English language, which constructs its entire literary cannon with only 26 letters, nature encodes the genomes of all of its great creatures with the same 20 amino acids. These 20 amino acids can conveniently be represented with one-letter codes: A for alanine, C for cystine, D for aspartic acid, etc. If a genome is analogous to a book, and an amino acid is analogous to a letter, a paragraph would likely correspond to a protein, and a chapter to a group of related proteins, known as an operon.

To define the concept of a genomic word, a basic element that has meaning for many organisms consider Fig. 1, which shows how the same

English	Science is the great antidote to the poison of superstition.	sentence from different languages share n-grams that can be used to associate meaning with words. This observation has been exploited by Google translate. For biology, shared k-mers of amino acid play the same role. Figure 2 shows a histogram of the number of amino acid exact
French	La Science est le grand antidote au poison de la superstition.	
Spanish	La ciencia es el gran antídoto al veneno de la superstición.	
Dutch	Wetenschap is het grote tegengif aan het vergift van bijgeloof.	
German	Wissenschaft ist das grosse Antidot zum Gift von Aberglauben.	

Fig 1. The same sentence from different languages share n-grams that can be used to associate meaning to words.

matches between *E. coli* and a variety of other bacteria as a function of the length of this exact match. Short patterns of amino acids are ubiquitous, with 300 million instances of matching amino acid 3-mers

between the divergent organisms *E. coli* and *B. subtilis*. They are too numerous to help convey meaning. On the other hand, exact matches of 20-mers of amino acids occur mainly in groups of closely related organisms. They are useful in identifying organisms in that group, but cannot be used to learn about common functionalities.

There are typically 2,000 matching k-mers of length 10 between divergent organisms and over tens of thousands shared 10-mers for more closely related organisms. As a comparison, we expect one random match between any two genomes. Independently, David Baker identified 9-mers of amino acids as a relevant unit for protein structure prediction (rosetta). These shared 10-mers arise both through natural selection and inheritance and are thus expected to be present in organisms not in our database [2].

The recently published paper by the authors [2] exploits this idea and combines it with evolutionary theory and web search-engine technology to develop a software package Sequedex [3] that can classify raw metagenomics reads 250,000 times faster than the current pipelines.

This new tool makes possible a much more novel use of next-generation sequencing, more along the lines proposed by the National Academy study—that of identifying microbial communities through their phylogenetic profile. Figure 3 shows the rolled-up phylogeny of 242 environmental microbial communities obtained from 30 separate studies and obtained from National Center for Biotechnology Information Sequence Read Archive and the Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis website. Literature

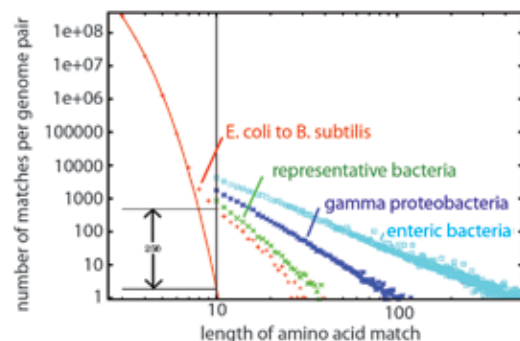


Fig. 2. Graph of the number of shared  $k$ -mers between *E. coli* and *B. Subtilis* as a function of  $k$ , on a log-log scale. Initially, the number of  $k$ -mers decrease exponentially. At  $k=8$ , the exponential decrease becomes polynomial (powerlaw). Extrapolation of the exponential decay to  $k=10$  shows that the signal to noise ratio of 10-mers is 250.

variable portion of the ribosome.

The Sequedex algorithm can be adapted to design viral and bacterial pan-diagnostics using shotgun sequencing. Figure 4 shows a histogram of viral reads identified from 100 million reads sequenced from an RNA sample preparation from a clinical diarrheal sample. The y-axis shows the number of distinct reads identified, plotted against the node-number of the RNA virus portion of a one-per-species tree of viruses. Four distinct peaks show up against an essentially flat baseline. Insets show the viral phylogeny in the region where matching reads were identified,

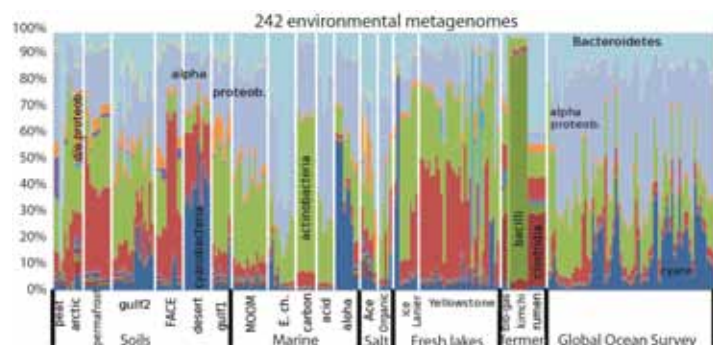


Fig. 3. Phylum-level phylogenetic profiles across 242 environmental metagenomic sample sets taken from 30 separate studies. Not only are distinctive profiles observable for the different sites, but field replicates from similar studies show a 99.9% similarity when comparing the full-resolution phylogenetic profiles to one another [2].

including adeno-associated virus, respiratory syncytial virus (RSV), sapovirus, and human enterovirus. The analysis, using Sequedex, took less than 15 minutes on a laptop. Competing methods [such as Basic Local Alignment Search Tool (BLAST)] applied to

references for each of the samples are available in the online documentation for the Sequedex software package [3]. The community similarities computed from signature-peptide-based distance metrics, such as that shown in Fig. 2, are much more stable and predictive than those determined by ribosomal RNA (rRNA) surveys because they are composed of the occurrence of thousands of conserved elements, rather than the

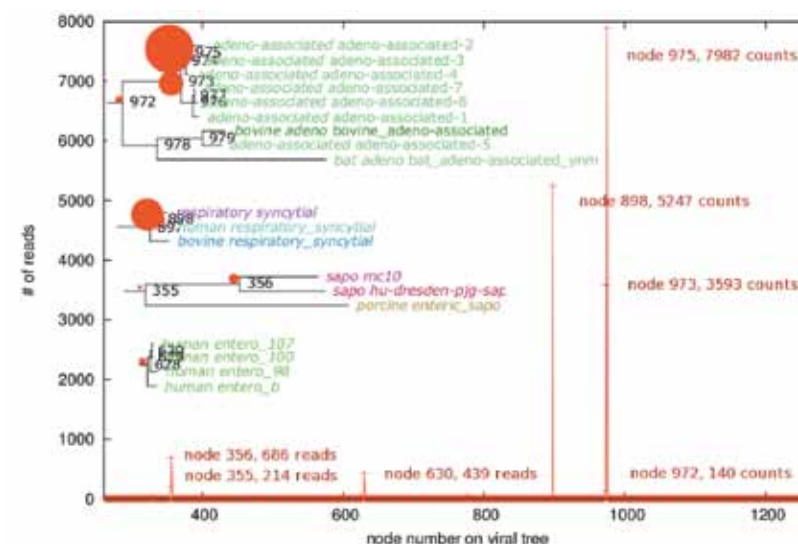


Fig. 4. Histogram of 100-bp reads identified in an RNA sequencing run of 100 million human, bacterial, and phage reads sequenced from 200-mL clinical diarrheal samples. Viral reads were identified in one hour of CPU time on a desktop computer, requiring 2 Gb of memory. Data were obtained from the laboratory of Charles Chiu, UCSF.

amino-acid translations and compared to the non redundant (NR) protein database require approximately one week of computing time on a cluster of computers, making the process unsuitable for routine use. As already noted, the process does not rely on the virus being closely related to a virus in the reference database, only that some of the genes contain conserved regions. Even such distinct pathogens as measles, mumps, and Nipah virus share signature peptides in numerous places across their genomes, which opens the possibility for broad characterization of endemic pathogens in a given environment.

[1] Handelsman J., et al., "The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet," National Research Council, Washington, DC (2007).

[2] Berendzen, J.R. et al., *BMC Res Rep* 5, 460 (2012).

[3] Sequex, <http://sequedex.lanl.gov> (2012).

#### Bibliography

Rohl, et al., *Meth Enzymol* 383, 66 (2004).  
Garten, et al., *Science* 325, 197 (2009).  
Korber, et al., *Science* 288, 1789 (2000).  
Ou, et al, *Science* 256, 1165 (1992). <http://hfv.lanl.gov>



# Time-of-Infection Estimated by a Time-Continuous HIV-1 IgG Model

Helena Skar,  
Thomas Leitner, T-6

The incidence of an infectious disease is vital for informed and targeted prevention, and knowing the time-of-infection is important for estimating the incidence in a population. For acute infections, like influenza, it is relatively straightforward to estimate time-of-infection because it occurred just shortly before the diagnosis. For chronic infections, like that of Human Immunodeficiency Virus type 1 (HIV-1), time-of-infection estimation is more complicated because most diagnosed persons have an established HIV-1 infection of unknown duration. In the last few years, there has been considerable interest in the development of biomarker assays that can determine if an HIV-1 infection is recent in order to estimate the incidence in a population. These serological assays are based on the knowledge about the development and maturation of the HIV-1 antibody response in infected persons. Until now, these assays have provided a binary result, that is, recent versus long-term infection, rather than a quantitative estimation of time since infection.

In this project, we developed a model describing the production of a specific type of IgG antibodies that can be detected by the IgG capture BED enzyme immunoassay (BED assay). The assay name “BED” signifies that it is based on a trimeric branched peptide with each branch derived from the immunodominant region of the gp41 glycoprotein of HIV-1 subtype B, circulating recombinant form (CRF) 01\_AE or subtype D.

Similar to many biological systems where the rate of reproduction is proportional to the existing population and limited resources, the growth of BED-specific IgG (BED IgG) following infection can be modeled by a logistic function. In order to account for patient variability of the BED IgG growth following HIV-1 infection, we developed a time-continuous mixed-effect logistic model describing the production of BED IgG. The model was trained on a large cohort metadata set from a previously published study [1] that included 2,927 longitudinal BED assay results from 756 patients. This model was then validated on a second large dataset, representing cross-sectional BED data from 819 Swedish patients newly diagnosed with HIV-1 infection between 2002 and 2010.

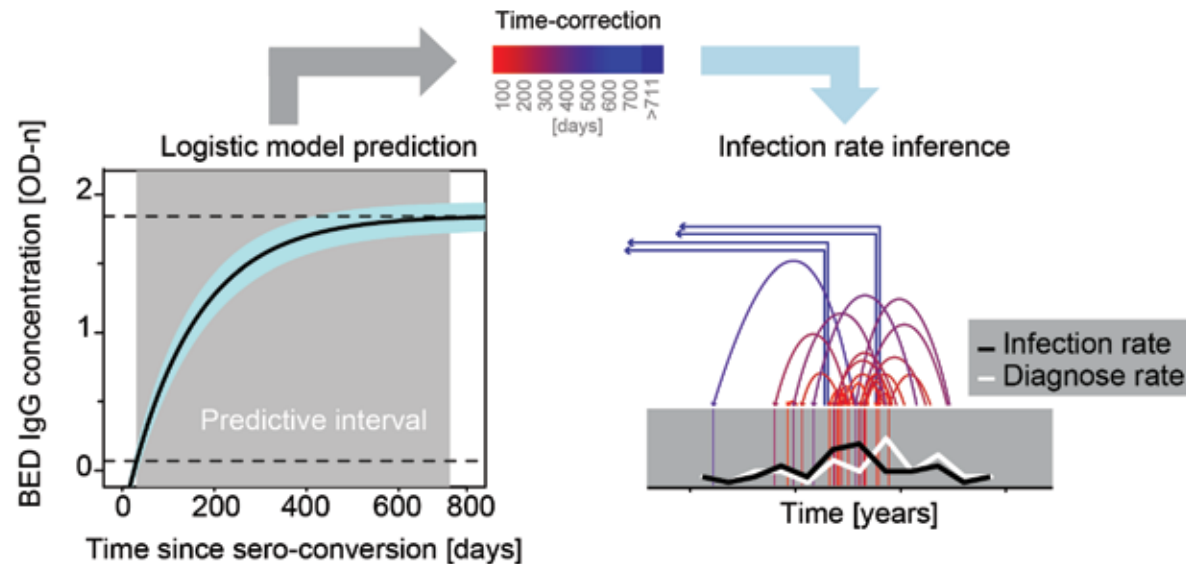
The validation data showed that the best parameterization of the logistic IgG model, which supports the biological intuition that there is no patient variation in BED IgG concentration (measured in normalized optical density, OD-n) at  $t=0$ . Thus, the typical patient is represented by a logistic growth of the BED-detected HIV-specific IgG following infection (Fig. 1). The model is informative of time since infection when the BED assay negative control is within an acceptable range, corresponding to  $\hat{t} = (0, 52)$  days, but specific to each batch of BED measurements, and

99% of the asymptote. The informative BED IgG OD-n interval translates into a continuous time interval with predictive power in  $t = (31, 711)$  days since seroconversion to when BED testing was done. Our model estimated the growth rate at  $r=0.00672$  OD-n units per day, and the asymptote at  $K=1.85$  OD-n units.

We next investigated whether the inferred time since seroconversion of our validation data could be described by the serological interval—that is, the time interval constrained by the last negative and first positive HIV-1 testing. Naturally, barring the time from infection to detectable HIV-1 by a valid method, infection must have happened sometime between these dates. The validation data was not from a cohort study, but rather data from patients detected by regular public health diagnostic testing. However, for this type of data it is not obvious that the population average time-of-infection is in the middle of the serological interval. Indeed, the non-cohort validation data shows a clear bias of infection-time shifted towards the date of diagnosis. Thus, this shows that (1) BED test results are applicable to infer the time since seroconversion in non-cohort type data, but (2) estimating date of infection as the midpoint between the last negative and the first positive HIV-1 test result is inaccurate and misleading in this type of data.

In conclusion, we have created a model that quantifies the time since seroconversion based on a simple serological assay, the BED assay. While we informed the model with BED assay results, because it is currently the most used biomarker for recency estimation, our model could easily be adjusted to other, future biomarkers. Our model is

applicable to biomarker results from patients included in cohort studies as well as patients diagnosed as a result of public health services. We expect that our method can improve incidence estimates, and thus provide valuable information for HIV-1 surveillance and prevention.



*Fig. 1. Graphical abstract. Lower left: logistic modeling of IgG-capture BED-enzyme immunoassay absorbance as a function of recency period. The resulting logistic model is predictive when BED OD-n = (0.07, 1.84), corresponding to recency periods of 31–711 days. This mixed-effects model describes the typical patient, where parameter values correspond to the whole population. Upper middle: correction-time (time between estimated time of infection and time of diagnosis) color legend. Lower right: Example from the Swedish data. The arrows show length of time-correction starting at the time of diagnosis and pointing at the inferred infection time, colored according to length of correction time. Lower part shows the inferred rate of diagnosis (grey line) and rate of infection (black line) for each quarterly year.*

# Insights into Microscopic Diffusion Processes at a Solid/Fluid Interface under Supercritical Conditions: A Study of the Aqueous Calcite ( $10\bar{1}4$ ) Surface

Chun-Yaung (Albert) Lu,  
Danny Perez, T-1;  
Donald D. Hickmott, TT-DO;  
Arthur F. Voter, T-1

We use molecular dynamics (MD) and non-Markovian kinetic Monte Carlo (KMC) simulations to investigate diffusion processes at an aqueous calcite ( $10\bar{1}4$ ) interface under various geologically relevant supercritical conditions. We show that solvent properties can change dramatically as a function of temperature and pressure and that complex interactions among the adsorbents, surface, and solvent molecules control the dynamics at the interface. Despite its apparent simplicity, this interface exhibits a non-trivial dynamical behavior. Our work proposes a new strategy to study fundamental issues related to solid-liquid interfaces at high temperatures and pressures.

Phenomena occurring at mineral interfaces in supercritical fluids are of fundamental importance in a wide range of Earth systems, including formation of ore deposits, ocean-floor hydrothermal systems, metamorphism within the continental lithosphere, geologic carbon sequestration, and geothermal energy production. Computer simulations are a promising tool for the study of such systems, especially considering the difficulty of carrying out experiments in such extreme conditions.

For example, the diffusion of adsorbed particles at a solid/fluid interface is a phenomenon of great fundamental interest and technological importance. It is known that adsorbed particles at a solid surface can undergo two types of surface translational motion—*in-surface self-diffusion* and *bulk-mediated surface diffusion* (BMSD) [1]. BMSD occurs when a fluid is present above the solid surface. An adsorbed particle can then desorb from the surface and re-adsorb elsewhere, leading to effective surface displacements. Since bulk diffusion is much faster than in-surface diffusion, lateral diffusion can be enhanced when bulk diffusion is involved.

In this work, the aqueous calcite ( $10\bar{1}4$ ) surface was chosen as a model system to study diffusion at solid/fluid interfaces. We use MD and KMC to simulate diffusion processes under various geologically relevant supercritical conditions (101.3 MPa, 300-800 K). A force field developed by Raiteri et al. [2] is used to simulate the interactions in the aqueous calcite system.

The free energy profiles of a calcium carbonate ( $\text{CaCO}_3$ ) molecule desorbed from the calcite ( $10\bar{1}4$ ) surface were first computed for  $T=300$ -800 K at 101.3 MPa. The results are summarized in Fig.1(a). Here the reaction coordinate is the vertical distance between the adsorbent carbon atom and the calcite surface. As  $T$  increases from

300 to 800 K, the bulk water density [Fig.1(b)] decreases from 1.030 to 0.477 g/cm<sup>3</sup>. The relative heights of the second and the third density peaks also decrease with  $T$ , indicating a decrease in ordering of the water at the interface. As a result, dissolution of  $\text{CaCO}_3$  becomes less favorable and the equilibrium position (corresponding to the free energy minimum) moves closer to the surface as  $T$  increases (Fig. 2).

One important consequence of the changes in the free energy profile is to modify the dependence of the desorption rate on  $T$ . Figure 1(c) shows that the ratio of the desorption free-energy barrier  $\Delta G_d$  to the thermal energy  $k_B T$  decreases from 300 to 600 K, so that in this region the tendency for desorption increases with  $T$  as we would expect. However, the trend reverses at 600 K, reflecting the sudden increase of the desorption free-energy barrier. This unusual desorption trend makes this an excellent model system for studying various diffusive regimes of BMSD.

We utilized a non-Markovian KMC algorithm with a set of predefined states (the set of free energy minima) and their corresponding first-passage time distributions to generate long state-to-state trajectories of  $\text{CaCO}_3$  monomer diffusion at the surface. The diffusion processes in the perpendicular (to the surface) and lateral directions are treated independently. Perpendicular diffusion is modeled by the Smoluchowski equation while the lateral diffusion is treated as free 2D Brownian motion with an effective diffusivity appropriate for that state.

The values of the mean lateral diffusivity,  $D$ , estimated from the long-time limit of the mean squared displacement (MSD) for various  $T$  and contact water thicknesses are summarized in Fig. 3(a). As seen in the figure,  $D$  increases with contact water thickness due to increasingly significant contributions from bulk diffusion.  $D$  also increases with  $T$  and reaches its maximum value at 600 K. For  $T > 600$  K, the system switches to inner-sphere type adsorption [Fig. 2(b)] and  $D$  consequently decreases.

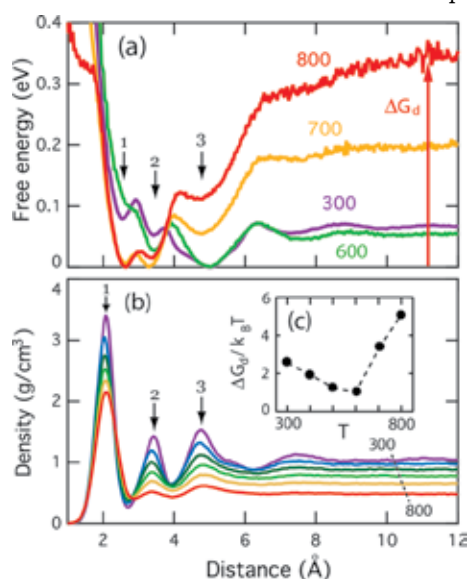


Fig 1. (a) Free energy of a  $\text{CaCO}_3$  molecule desorbed from the aqueous calcite ( $10\bar{1}4$ ) surface at various  $T$  and 101.3 MPa. (b) Mass density of water oxygen as a function of vertical distance. All densities were transformed to the corresponding water density. (c) The ratio  $\Delta G_d/k_B T$  as a function of  $T$ .



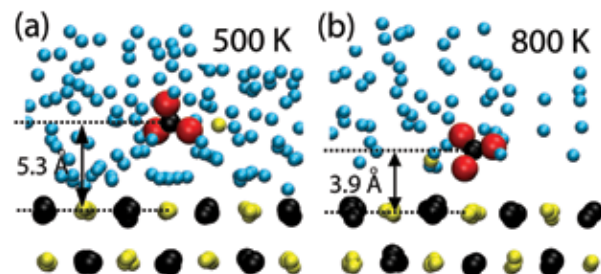


Fig. 2. Snapshots of MD simulations at 101.3 MPa at temperatures of (a) 500 K and (b) 800 K.

In order to validate the continuum model, we compare with direct MD simulations for a thickness of 12 Å [note black dots in Fig. 3(a)]. The agreement is very good, demonstrating that our model captures the essential factors controlling the diffusion process.

It is known that BMSD exhibits non-Fickian diffusion behaviors with anomalous scaling properties [1]. The short-time scaling of the diffusion process reveals more detailed information about the desorption kinetics. For an aqueous calcite system with infinite contact water, the short-time (~1 ns) effective MSD at various  $T$ s is summarized in Fig. 4.

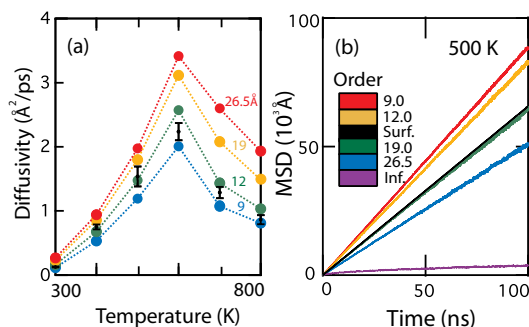


Fig. 3. (a) Monomer diffusivities obtained from the MSD curves for various temperatures and water thicknesses. The black dots with error bars are results from direct MD simulations with a contact water thickness of 12 Å. (b) Effective MSD as a function of time for various water thicknesses at 500 K. The MSD line estimated from surface diffusivity is also included in the plots (black solid lines).

By comparing the MSD trends with the in-surface diffusion lines (black), it is clear that the MSD curves at 300, 700, and 800 K exhibit a super-diffusive trend (i.e., they are concave and diffuse faster than in-surface diffusion) while others are sub-diffusive (i.e., they are convex and diffuse slower than in-surface diffusion).

From Fig. 4 it is clear that the desorption/adsorption kinetics affect the scaling properties of the effective surface diffusion. It has been reported that diffusion trends can also be affected by the contact fluid thickness [3]. We investigated this effect by varying the contact water thickness in the calcite system. For a given  $T$ , 500 K for example, the effective surface diffusion can be faster or slower than the in-surface diffusion depending on the water thickness [Fig. 3(b)]. It is interesting to note that the optimized water thickness that maximizes the effective surface diffusivity also varies with  $T$ .

In conclusion, we studied how the diffusion dynamics at a calcite/water interface change under different supercritical conditions. The MD and KMC simulations showed the effective lateral diffusion at the interface is greatly affected by the water conditions. This is a clear example of the tunable solvation power of supercritical fluids and of their effect on the diffusion dynamics of absorbed species. We believe these factors will affect the efficiency of surface mass transport and must be understood to predict reaction kinetics and transport efficiencies (including the relative importance of surface diffusion versus diffusion in bulk fluids) in conditions relevant to crustal geologic and energy systems.

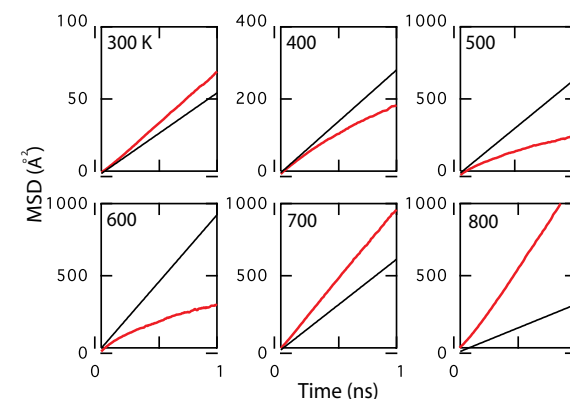


Fig. 4. Short-time effective MSD (red solid lines) as a function of time for various  $T$ . The calcite surface is in contact with infinite water. Black solid lines represent the same quantities for in-surface diffusion.

- [1] Bychuk, O.V. and B. O'Shaughnessy, *Phys Rev Lett* **74**, 1795 (1995).
- [2] Raiteri, P. and J.D. Gale, *J Am Chem Soc* **132**, 17623 (2010).
- [3] J. Revelli, A. et al., *Eur Phys J B* **37**, 205 (2004).

# Close Integration between Theory and Experiment: Simulations and Chemical Probing Experiments of Molecular Switches

Karissa Sanbonmatsu,  
Scott Hennelly, T-6;  
Jose Onuchic,  
Ryan Hayes, Rice University

A riboswitch is an RNA-based molecular switch that plays an important role in regulating bacterial metabolism. This switch senses the presence of a small molecule with high specificity and makes a go/no-go decision for gene expression based on the concentration of the small molecule. In the case of the S-adenosylmethionine (SAM) riboswitch, the riboswitch is embedded in the messenger RNA (mRNA) upstream of the coding region for the enzyme SAM synthetase, which synthesizes SAM. In absence of SAM, a terminator helix forms in the RNA, turning gene expression off and preventing the production of SAM synthetase (Fig. 1). In the presence of SAM, the RNA folds into a different structure that does not contain the terminator helix, allowing production of SAM synthetase. Thus, a single sequence can have two mutually exclusive structures. High concentrations of ligand shift the equilibrium towards the aptamer structure; low concentrations produce a competing structure called the expression platform structure. Using a combination of molecular dynamics (MD) simulations and biochemical probing experiments, we show that divalent ions play a key role in the mechanism of the SAM riboswitch.

RNA molecules are similar to proteins in the sense that they undergo an analogous folding process, participate in conformational changes and catalyze various chemical reactions. One important difference is that the folding and function of RNA molecules is almost always very sensitive to magnesium ions. This is largely a result of the highly charged backbone of the RNA molecule. The exact mechanism of magnesium action is not well understood at the atomistic level. In particular, each magnesium ion is thought to couple with RNA using

one of two interaction modes: (1) site-specific binding, or (2) participation in a diffuse magnesium cloud surrounding the RNA.

To elucidate the magnesium action mechanism, we first performed an extensive series of explicit solvent MD simulations of the SAM riboswitch with total sampling of 20  $\mu$ s [1]. Ten simulations of 2  $\mu$ s each were performed using a periodic box of water (length of

side = 100 Å). Simulations were performed with magnesium concentrations of 0 mM, 1 mM, and 10 mM. A key prediction resulting from this study is that one particular magnesium ion is chelated, interacting directly with the nonbridging phosphate oxygens of residues A10 and U64 in the X-ray structure solved by Batey and co-workers [2] (A10 and U71 in residue numbering of the native wild type SAM riboswitch) (Fig. 2). This was the

only magnesium ion site specifically bound to the RNA for the entire simulation.

To test this prediction, we used a biochemical technique called nucleotide analog interference mapping (NAIM). Here, riboswitch RNAs are synthesized that contain modifications at random residues (approximately one modified residue in each RNA molecule). In this set of experiments, phosphothioate modifications were incorporated, replacing a nonbridging phosphate oxygen with a sulfur, which significantly reduces the affinity of magnesium to the RNA in a site-specific manner. The NAIM method requires an assay with which one can measure the effect of the modification. We used an assay that measures switching between the aptamer and expression platform structures. In this assay, the riboswitch is divided into two pieces: the aptamer and a portion of the expression platform. Here, the expression platform strand is an RNA/DNA chimera of 2'-O-methyl RNA and DNA residues.

Formation of the expression platform helix creates sites for RNase H cleavage, allowing us to measure the amount of helix formation using RNase H digestion. In the experiments, the aptamer is first folded. Next, the expression platform strand is added. In the absence of magnesium, this strand easily invades the aptamer resulting in formation of the expression platform. In the presence of magnesium, the aptamer is stabilized, preventing formation of the expression platform helix. By comparing wild type and modified RNAs, we are able to determine specific residues, that, when modified, interfere with aptamer stabilization. Because the phosphothioate modification blocks site-specific magnesium binding, the assay reveals specific magnesium chelation sites on the RNA (Fig. 3). We find that this occurs at only four

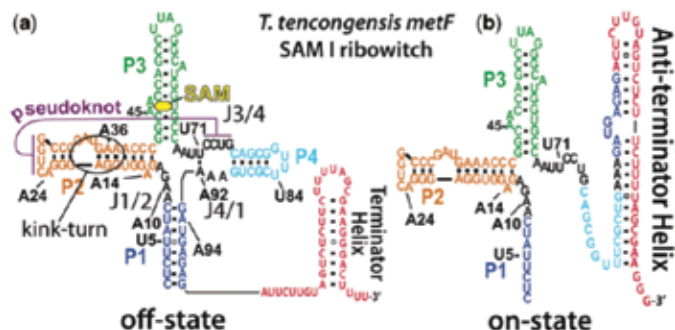


Fig. 1. Secondary structure of the SAM riboswitch RNA molecule. (a) Off-state is formed in the presence of SAM. Terminator helix prevents gene expression. (b) On-state is formed in the absence of SAM. The anti-terminator helix, located in the expression platform domain, precludes formation of the terminator, allowing gene expression.

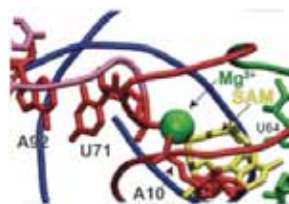


Fig. 2. MD simulations predict a site-specifically bound, chelated magnesium ion (green) connecting riboswitch residues A10 and U64 (A10 and U71 in residue numbering of the native wild type SAM riboswitch).

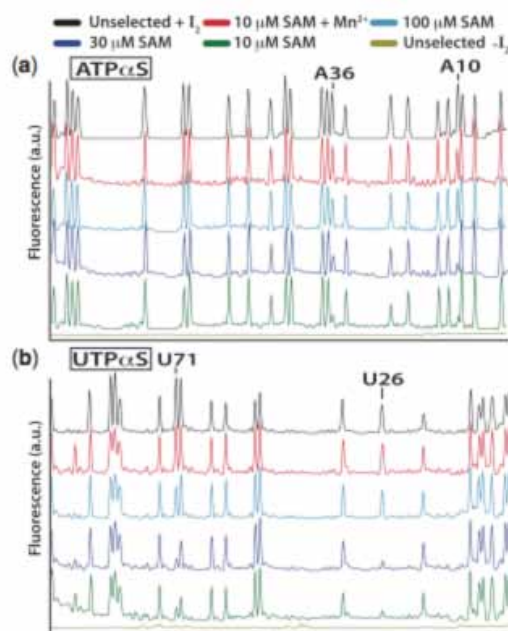


Fig. 3. NAIM shows that a chelated magnesium ion exists near riboswitch residues A10 and U64 (A10 and U71 in residue numbering of the native wild type SAM riboswitch).

sites, including two that we predicted with our simulations: A10 and U64 (A10 and U71 in residue numbering of the native wild type SAM riboswitch) [3].

In our simulations, increasing the magnesium concentration slows fluctuations in the RNA. This result is consistent with a set of chemical probing experiments we performed on the riboswitch [3]. Here, we used selective 2'-hydroxyl acylation by primer extension (SHAPE) to measure the backbone mobility of each residue under a wide variety of magnesium and SAM concentrations. We find that the addition of magnesium substantially reduces the RNA mobility.

Interestingly, our experimental study also shows that magnesium and SAM work together to control riboswitch function. In particular, for correct riboswitch operation the addition of SAM needs to produce the aptamer structure and preclude formation of the expression platform helix. However, we find that this does not occur in the presence of SAM alone, even at high concentrations of SAM. Addition of magnesium as well as SAM is required for formation of the fully collapsed aptamer structure. Similarly, magnesium alone cannot produce the fully collapsed aptamer. Some level of SAM is required in addition to the magnesium.

When we performed SAM and magnesium titrations in the context of our two-piece switch assay, we discovered that magnesium has the effect of maintaining switch polarity. That is, in the absence of SAM, higher concentrations of magnesium stabilize the expression platform helix; in the

presence of SAM, higher concentrations of magnesium stabilize the aptamer structure.

Rather than a few site-specifically bound ions surrounded by a continuum of diffuse ions, we observe a layer of coordinated magnesium that is transiently bound but strongly coupled to the RNA. After measuring diffusion of magnesium ions as a function of distance from the RNA, four categories of ions emerged: (1) free ions, unperturbed by the RNA; (2) diffuse ions with high diffusion coefficients; (3) site-specifically bound ions with near-zero diffusion; and (4) a class of ions called outer-sphere ions with intermediate levels of diffusion (Fig. 4). Surprisingly, outer-sphere ions comprised 80% of the magnesium ions in the simulation. These outer-sphere ions are energetically frustrated and exhibit glass-like behavior, hopping from site to site within the core of the RNA structure and on the surface of the RNA structure. They retain their inner-sphere waters and are separated from the RNA by a single hydration layer. We see that these ions fluctuate on time scales comparable to the fluctuations of the riboswitch and couple to the riboswitch. Their diffusion is two orders of magnitude lower than the bulk magnesium.

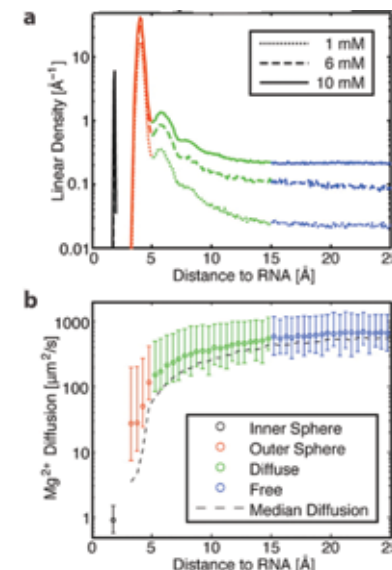


Fig. 4. Explicit solvent MD simulations (20  $\mu$ s total sampling) reveal four categories of magnesium ions: inner sphere (site-specifically bound), outer sphere, diffuse, and free. The outersphere ions constitute 80% of the magnesium ions. These ions are glass-like, hopping from site to site within the RNA due to energetic frustration.

[1] Hayes, R.L. et al., *J Am Chem Soc* **134** (29) 12043 (2012).

[2] Stoddard C.D. et al., *Structure* **18**(7) 787 (2010).

[3] Hennelly, S.P. Novikova, I.V., and Sanbonmatsu K.Y., "The Expression Platform and the Aptamer: Cooperativity between Mg<sup>2+</sup> and Ligand in the SAM-I Riboswitch," *Nucleic Acids Res*, in print (2012).



# Computational Study of Electrochemical Windows of Room-Temperature Ionic Liquids

Yong-Hui Tian, T-1;  
George S. Goff,  
Wolfgang H. Runde, C-IIAC;  
Enrique R. Batista, T-1

Room-temperature ionic liquids (RTIL) are regarded as green solvents due to their low volatility, low flammability, and thermal stability. RTILs exhibit wide electrochemical windows, and they show potential as media for a variety of electroplating processes. Therefore, understanding the factors dominating the electrochemical stability of RTILs is crucial for improving the efficiency of separation processes. In this study, we computationally investigated the electrochemical properties of a series of imidazolium-based RTILs. The general trend of electrochemical windows of the RTILs studied is well reproduced, showing that it increases in the order of imidazolium < ammonium < pyrrolidinium < phosphonium giving confidence to the methodology presented to use it in screening studies of ionic liquids [1].

Room temperature ionic liquids (RTIL) are salts that are in molten states, typically below 100°C. Unlike common inorganic salts, which are composed of small-size ions, RTILs are comprised of bulky organic cations and ions. Due to their liquid state at room temperature, RTILs have found applications as solvents in various chemical and electrochemical processes. In comparison to conventional organic solvents, RTILs have the advantages of negligible vapor pressure, high thermal stability, low flammability, and tunability by fictionalizing the chemical structure of ions. Due to these advantages, they are considered

to be environmentally friendly solvents. Another notable property of RTILs is their intrinsic electrical conductivity, making them useful as solvents in electrochemical systems without the need for supporting electrolytes. In recent years, RTILs have been used in a variety of electrochemical processes such as lithium ion batteries, fuel cells, solar cells, and electro-depositions.

From the viewpoint of practical electrochemical operations, the electrochemical stability

of electrolytic media is one of the most important factors determining the utility and performance of electrochemical processes or devices. The electrochemical stability of the ionic liquid is characterized by the electrochemical window (ECW). The electrochemical window is defined as the potential interval between the cathodic and anodic potential

limits at which reduction and oxidation reactions of electrolytes occur at the electrodes. In this project, we aim to find reliable and efficient methods for modeling the electrochemical windows of the commonly used RTILs. With such methods, we hope, in future work, to screen for suitable candidate RTILs for certain electrochemical purposes among the huge number of RTILs developed so far, and to make predictions of the electrochemical properties for new RTIL systems.

We proposed four different approaches including gas phase method, polarizable continuum model (PCM), quantum mechanics/molecular mechanics (QM/MM), and thermodynamic cycle (TDC) method by running classical molecular dynamics (MD) simulations. These methods cover different degrees of approximation and computational cost from gas-phase calculations to full explicit solvation models. In the PCM model, as shown in Fig. 1(a), the solvent was treated as a continuous medium with the dielectric constant of the ionic liquid. An individual ion was embedded in solvents and a cavity was created, the surface of which was polarized to mimic the polarization of the solute species. For the QM/MM method, the reaction free energy changes were calculated using the two-layer ONIOM method of Morokuma and coworkers [2]. In this method, the reactive species are treated by first principle, while the environment is described by MM as illustrated by Fig. 1(b). The structures used for ONIOM calculations were taken from pre-equilibrated configurations with classical MD simulations. For the thermodynamic cycle method as depicted by Fig. 1(c), the calculation of free-energy change is reduced to two parts, the reaction free energy in the gas phase was calculated quantum mechanically, and the solvation free energy of the reactive species was estimated with thermodynamic integration. The modeling was based on the ionic liquid systems composed of imidazolium, ammonium, and phosphonium cations, and bistriflimide anions.

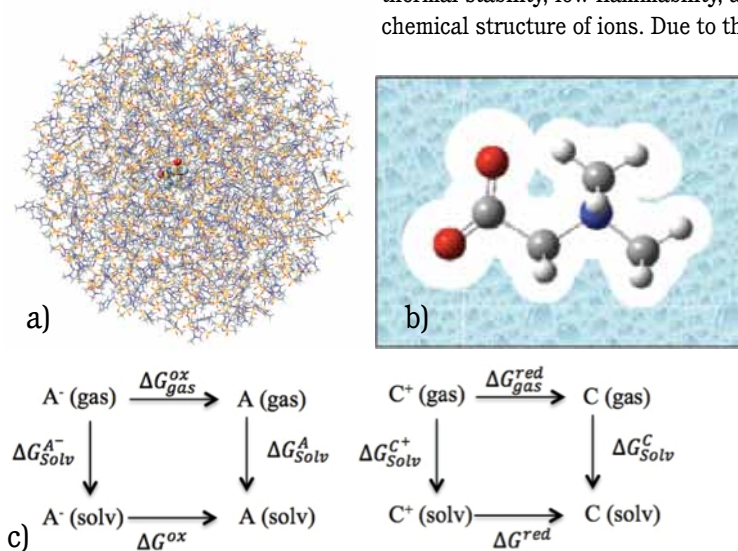


Fig. 1. Schematic diagram of the computational methods: (a) QM/MM; (b) PCM; (c) thermodynamic cycles (TDC).

As shown in Table 1, the calculated windows via the gas phase method are far from experimental values. However, the calculated windows via the isolated-ion method PCM, QM/MM, and TDC methods predict consistent results that agree fairly well with experiment for the ionic system of [C<sub>2</sub>MIM][Tf<sub>2</sub>N]. Therefore, solvation effect plays an important role in determining RTIL electrochemical windows. We found that to predict the electrochemical windows of RTILs, the PCM model represents a reasonable compromise between accuracy and efficiency. Figure 2 shows the PCM calculated electrochemical windows against the experimental values for a set of RTILs. The general trend reproduced well, showing that the windows increase in the order of imidazolium < ammonium < pyrrolidinium < phosphonium.

**Table 1.** The calculated potentials and electrochemical windows (in V) of [C<sub>2</sub>MIM<sup>+</sup>][Tf<sub>2</sub>N<sup>-</sup>]

	$V_{ox}$	$V_{red}$	ECW	Exp.
<i>vacuo</i>	5.40	-4.56	0.84	4.1
PCM	7.00	-2.72	4.28	4.5
TDC	7.02	-2.55	4.47	
ONIOM	7.55	-2.94	4.61	

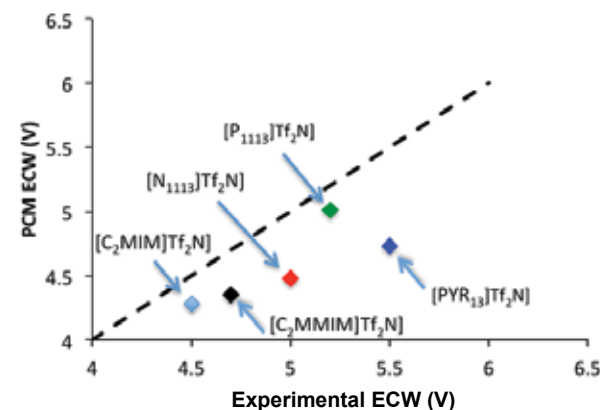
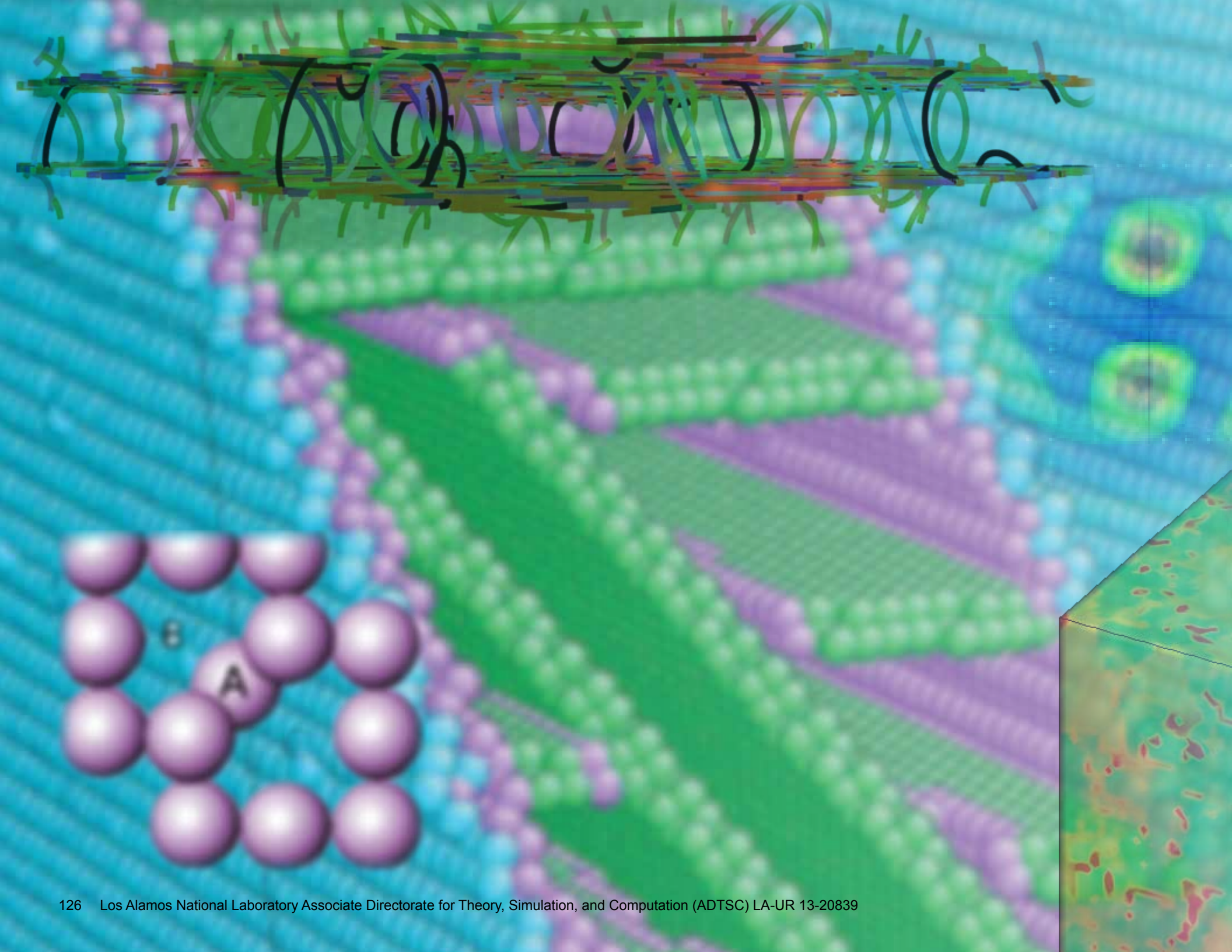


Fig. 2. The plot of experimental ECW against the PCM calculations.

[1] Tian, Y.-H. et al., *J Phys Chem B* **116**, 11943 (2012).

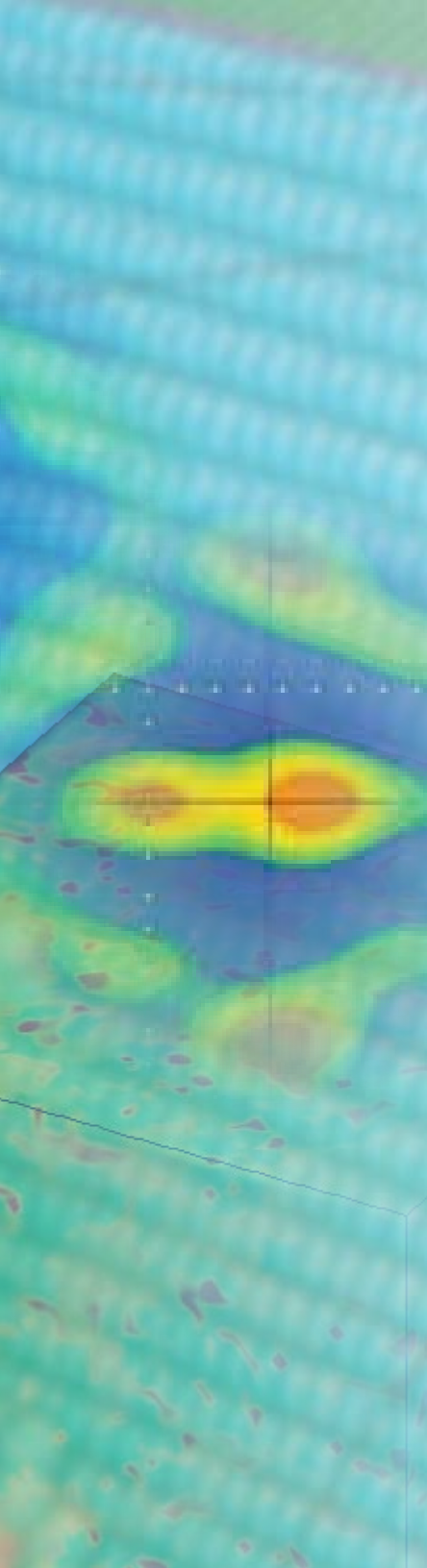
[2] Dapprich, S. et al., *J Molecular Structure: THEOCHEM*, **1**, 461 (1999).







# Materials Science



Researchers at LANL conduct fundamental research into the state and nature of materials, provide new theories for materials behavior, and create and provide new constitutive properties models and parameters for a wide variety of simulation tools in use or under development around the world. Not surprisingly, much of our work deals with materials in extreme conditions—highly stressed, shocked, and high-radiation field environments. This work with materials in extreme conditions in particular supports the LANL grand vision to design and construct a world-class facility called MaRIE (Matter-Radiation Interactions in Extremes) to predict and control material performance under dynamic and harsh conditions. Articles in this section explore methods to increase the performance of electronic structure and quantum-based molecular dynamics simulations, coupling a thermal mechanical model with a thermal ignition model for energetic materials, predicting the evolution of crystallographic texture for rolled metals, discovering the design principles for large spin and phonon fluctuations, calculating transition rates in large-scale materials, modeling dislocation processes of nanolayered composites under shock compression, capturing the fundamental physics of dislocation-interface interactions, and simulating dynamics and pattern formation of driven dislocation assemblies. This broad range of physical modeling approaches, mathematical and numerical algorithms, and multiple length and time scales is illustrative of the breadth and depth of our capabilities in this arena.

# Quantum-Based Molecular Dynamics on Graphics Processing Units

Susan M. Mniszewski, CCS-3;  
Anders Niklasson,  
Ed Sanville,  
Marc J. Cawkwell, T-1

The performance of quantum-based molecular dynamics (MD) simulations using the LANL-developed code LATTE is limited by the time required to compute the density matrix at each time step. The density matrix has been computed historically via the diagonalization of the Hamiltonian matrix. We have investigated the performance and accuracy of the second-order spectral projection (SP2) algorithm that calculates the density matrix via a recursive series of generalized matrix-matrix multiplications. Owing to its simplicity, the SP2 algorithm is ideally suited to implementation on graphics processing units (GPU). We demonstrate that optimized implementations of this algorithm on single and multiple GPUs on a single compute node lead to significant speed-ups with respect to central processing unit-based implementations and diagonalization. Furthermore, the SP2 algorithm produced an error rate in the density matrix that are as good or better than traditional methods.

MD is a popular and widely used simulation technique for the study of the evolution of a collection of atoms over time [1]. MD simulations require an interatomic potential that gives the potential energy of the system and the forces acting on each atom as a function of the relative coordinates of all of the atoms [2]. Interatomic potentials that describe explicitly the electronic structure of molecules and solids are the most physically accurate, yet their application in MD simulations has

been severely limited by their prohibitive computational expense [3]. As a result, there is a constant demand for the development of better algorithms and computational methods for the acceleration of quantum-based interatomic potentials such that ever more challenging simulations become tractable.

Modern general purpose GPUs are attractive candidates for the acceleration of many compute-intensive applications on account of their very high memory bandwidth and peak number of floating point operations per second (FLOPs) relative to CPUs. However, GPUs are most efficient only when high levels of thread-level parallelism can be

extracted from the algorithm, branching within warps can be avoided, and CPU-GPU communication can be minimized. Thus, not all algorithms are amenable to implementation on GPUs.

The density matrix,  $\rho$ , is a key quantity in the computation of the potential energy and interatomic forces in many quantum-based

potentials. In dense matrix algebra, the time required for the computation of the  $M \times M$  density matrix scales with the cube of the matrix dimension,  $O(M^3)$  [3]. As a result, the calculation of the density matrix may dominate the total computational time, particularly for large systems. In the quantum-based MD code LATTE we have pursued an algorithm for the computation of the density matrix that avoids many of the intrinsic limitations of GPUs.

The SP2 algorithm [4,5] enables the density matrix to be computed directly from the Hamiltonian via an expansion of the Fermi operator,

$$\rho = \Theta[\mu I - H] = \lim_{i \rightarrow \infty} f_i[f_{i-1}[\dots f_0[X_0]\dots]]$$

where  $\Theta$  is the matrix Heaviside step function,  $H$  the Hamiltonian matrix,  $\mu$  the chemical potential,  $I$  the identity matrix.  $X_0$  is equal to the Hamiltonian matrix rescaled such that its eigenvalues occupy in reverse order the interval  $[0,1]$ , and

$$f_i[X_i] = \begin{cases} X_i^2 & \text{if } 2\text{Tr}[X_i] \geq N_e \\ 2X_i - X_i^2 & \text{if } 2\text{Tr}[X_i] < N_e \end{cases}$$

where  $\text{Tr}$  denotes the matrix trace and  $N_e$  is the total number of electrons. The value of the chemical potential is adjusted automatically in the SP2 algorithm to obtain the correct occupancy.

The SP2 algorithm was ported to Nvidia GPUs using CUDA version 4.2 with Nvidia's CuBLAS implementation of the level 3 BLAS DGEMM subroutine for performing the generalized matrix-matrix multiplications,  $C \leftarrow \alpha AB + \beta C$ , where  $A$ ,  $B$ , and  $C$  are matrices and  $\alpha$  and  $\beta$  are scalars.

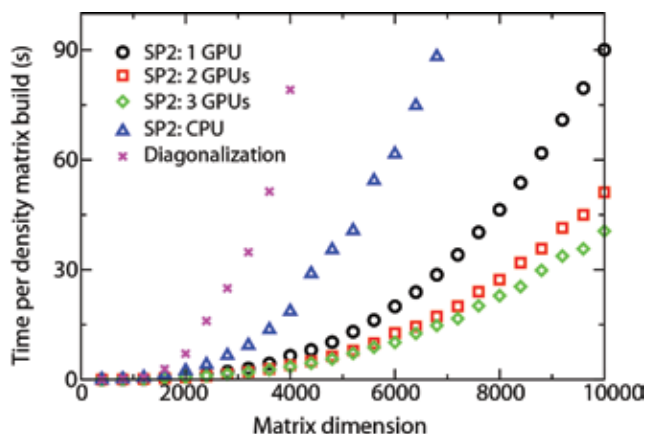


Fig. 1. Time per density matrix construction with LATTE via diagonalization and the SP2 algorithm on two hex-core Xeon processors and on multiple Nvidia C2090 GPUs on the Keeneland cluster at the National Institute for Computational Sciences (NICS).



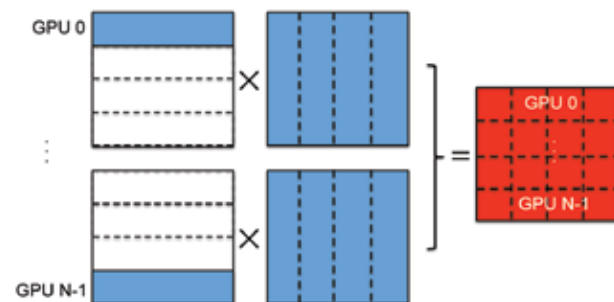


Fig. 2. Schematic illustration of the partitioning of sub-block matrix-matrix multiplications among  $N$  GPUs. Each GPU has a copy of the entire symmetric matrix.

converged [5].

In Fig. 1 we present a series of timings for the computation of the density matrix for liquid methane in LATTE [6] using diagonalization and the SP2 algorithm on two hex-core CPUs versus our GPU implementations of the SP2 algorithm. Here it is evident that the performance of the implementation on one GPU exceeds those of an optimized CPU implementation and diagonalization even for relatively small systems [5]. Furthermore, an analysis of the errors in the final density matrices shows that the SP2 algorithm is as accurate or better than traditional diagonalization [5].

Since multiple GPUs are commonly installed on a single compute node, we have investigated a parallel GPU implementation of the SP2 algorithm. We designed and implemented a simple scheme for the parallelization of the generalized matrix-matrix multiplication that is illustrated schematically in Fig. 2. Here, the sub-blocks into which the matrices  $X_i$  are divided were purposefully kept as large as possible owing to the problem-size dependence of the FLOP rate of GPUs. The CuBLAS DGEMM was used to perform all of the sub-block matrix-matrix multiplications. The matrix traces and matrix-matrix additions were also performed in parallel across multiple GPUs. The timings presented in Fig. 3 show clearly that the use of multiple GPUs results in improvements in performance as the dimensions of the matrices increase. The overheads encountered for small matrices arise from the problem-size dependence of the FLOP rate of the GPUs and communication between GPUs via the CPU. Nevertheless, our results show a convergence

We were able to minimize expensive communication between the CPU and GPU by porting the entire subroutine to the GPU such that we transfer only the matrix  $X_0$  to the GPU at the start of the algorithm and pull back  $\rho$  once the recursive expansion has

toward ideal speed-ups as the matrix sizes increase with no loss of accuracy with respect to the CPU code or the implementation on one GPU.

In conclusion, hybrid multi-core CPU and multi-GPU computational architectures have been employed to significantly increase the performance of electronic structure and quantum-based MD simulations in LATTE. GPUs are particularly attractive for use as accelerators, especially when one considers their performance on a per Watt or per unit cost basis. However, users will enjoy performance gains only for those algorithms that are well tailored via computational co-design toward the intrinsic strengths of GPUs.

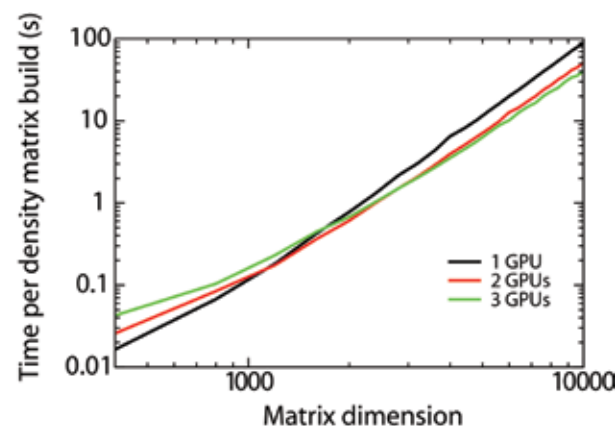


Fig. 3. Relative performance of the SP2 algorithm on one, two, and three Nvidia C2090 GPUs as a function of the problem size. Running on one GPU is fastest for small problems while three GPUs provide best performance when the dimensions of the density matrix exceed about  $3000 \times 3000$ .

- [1] Allen, M.P. and D.J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press (1987).
- [2] Finnis, M.W., *Interatomic Forces in Condensed Matter*, Oxford University Press (2003).
- [3] Goedecker, S., *Rev Mod Phys* **71**, 1085 (1999).
- [4] Niklasson, A.M.N., *Phys Rev B* **66**, 155115 (2002).
- [5] Cawkwell, M.J. et al., *J Chem.Theory Comput*, **8**, 4094 (2012).
- [6] Sanville, E.J. et al., "LATTE," LA-CC 10-004; <http://savannah.nongnu.org/projects/latte> (2010).



# Thermally Induced Damage in Energetic Materials

Bradford E. Clements,  
Axinte Ionita, T-1

The ability to correlate thermally induced damage in energetic materials to chemical ignition thresholds is an important goal for several DOE programs, including the Advanced Simulation and Computing Program (ASC), Joint Munition Technology Development Program (JMP), and C-2. We are addressing this problem by using both continuum and mesoscopic models and simulations, as well as drawing upon LANL's powerful experimental capabilities.

**E**nergetic materials research at LANL on plastic-bonded explosives has a long history dating back to nearly the inception of the Laboratory. Much of the research tended to focus on several key areas including equation of state (EOS), mechanical properties, chemical ignition, and detonation propagation. Mechanical properties were studied primarily to gain confidence on explosive safety handling. As a result, operating temperatures for mechanical experiments were typically restricted to the vicinity of room temperature. A high-temperature measurement might reach 50 to 70°C, for example. These temperatures are well below that of thermal ignition for an explosive like PBX 9501, a primary DOE explosive. Only recently has considerable interest shifted to investigating the mechanical properties of explosives at much higher temperatures. The reason for the shift in interest is the need to correlate mechanical properties, such as damage to thermal ignition properties, which for PBX 9501 requires temperatures in the range of 160 to 200°C. Correlating mechanical properties to ignition is itself important because of the need to predict thermal ignition thresholds of an explosive in various low-velocity impact scenarios. In a low-velocity impact, the mechanical properties of the explosive have a dominating influence on the outcome of an initiated chemical reaction. The mechanical properties will determine if the reaction extinguishes, builds up to a full detonation, or results in something in between. To model such events requires coupling a thermal mechanical model with a thermal ignition model, which is the ultimate goal of the work described here.

Rae, Parker, and Dickson [1] have shown that high temperatures have a profound influence on the mechanical properties of PBX 9501. They measured the stress-strain behavior at room temperature and then again

near 200°C, which is near the  $\beta$ - $\delta$  phase transformation. The  $\beta$ - $\delta$  phase transformation is a solid-solid transition occurring in the energetic component of PBX 9501, called HMX, and is accompanied by a 7% volume increase. Figure 1 shows several of their measured stress-strain curves where it is clear that a mechanism change occurs in going from room to high temperature. This mechanism change is potentially linked to the  $\beta$ - $\delta$  transformation, but other mechanisms may also contribute to the observed change in behavior. Interestingly, each proposed mechanism increases the level of damage in the explosive. Because of a large coefficient of thermal expansion mismatch between the PBX constituents, elevating the temperature results in a large increase in potential nucleation sites for damage. This will be exacerbated by the  $\beta$ - $\delta$  transformation. Upon straining the explosive, as in Fig. 1, these nucleation sites grow in size into fully macroscopic flaws that influence the mechanical stress-strain behavior.

Figure 2 is the solution of a LANL-developed continuum thermo-mechanical model. The model has the physics of a viscoelastic material that damages by the growth of shear cracks and porosity. It is clear that the same trend observed in the experiment is captured by the model if one lets the number density of nucleation sites increase with temperature, as expected. This is the first model known to capture this effect. The model relies heavily on measured thermo-mechanical properties of the explosive, and many of these properties were obtained by LANL's experimental groups. Nevertheless, for certain properties, macroscopic measurements are difficult. One such property is the microscopic interfacial debonding between the HMX and the polymer-binding matrix that occurs at elevated temperatures. To address this problem, we invoke a Direct Numerical Simulation (DNS) approach. DNS on PBX 9501 attempts to use micron-level radiography scans of the PBX 9501, obtained by B. Patterson at LANL, to obtain the microstructure. Using this microstructure and the properties of the explosive constituents, a finite-element simulation is then used to determine the

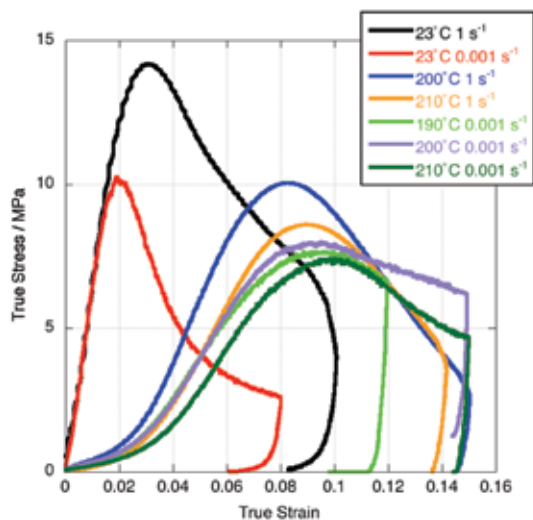


Fig. 1. Stress-strain curves for PBX-9501 measured by Rae, Parker, and Dickson [1].

underlying stresses, strains, and damage that occur in the explosive. The left part of Fig. 3 shows the DNS cell, where the lighter shaded areas are the HMX grains and the darker areas are either polymer binder, or a mixture of small HMX grains and binder. The right figure shows the thermal strains resulting when the temperature is increased by 100°C. The dark red regions are the new sites for damage nucleation. Using DNS we hope to provide quantitative input to the continuum model.

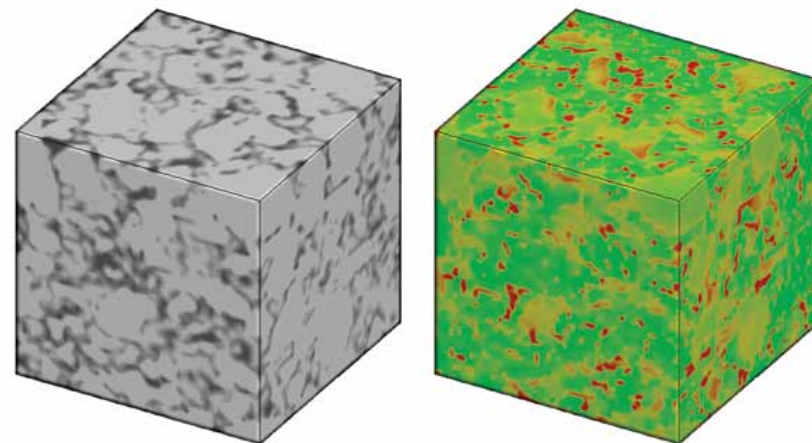
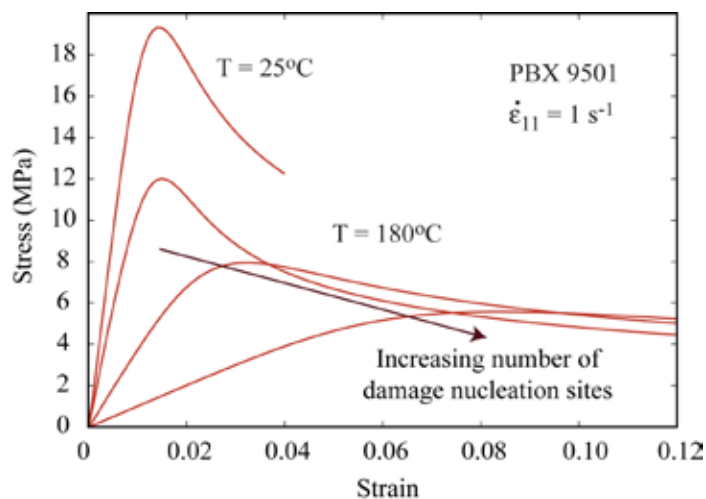


Fig. 3. A DNS representative volume element of PBX 9501 (left), and the corresponding thermal simulation (right), where the temperature has been increased by 100°C. The red regions have undergone substantial thermal straining. The box length is 750  $\mu\text{m}$ .

Fig. 2. Solution of our continuum model used for a study of effects of varying the damage nucleation site number density.



# Modeling the Texture Evolution of Cu/Nb Layered Composites During Rolling

Benjamin L. Hansen, T-3;  
John S. Carpenter, MST-6;  
Stephen D. Sintay, IAT-3;  
Curt A. Bronkhorst, T-3;  
Rodney J. McCabe, MST-6;  
Jason R. Mayeur,  
Hashem M. Mourad,  
Irene J. Beyerlein, T-3;  
Nathan A. Mara, Center for  
Integrated Nanotechnologies;  
Shuh-Rong Chen,  
George T. Gray III, MST-8

Metallic-based multi-layered nano-composites are recognized for their increased plastic flow strength, increased ductility, improved radiation damage resistance, improved electrical and magnetic properties, and enhanced fatigue-failure resistance compared to conventional metallic materials. One of the ways in which these classes of materials are manufactured is through accumulated roll bonding where the material is produced by several rolling and heat treatment steps during which the layer thickness is reduced through severe plastic deformation. In this article, a single rolling pass of the accumulated roll bonding process in which a Cu/Nb layered composite with an initial average layer thickness of 24  $\mu\text{m}$  subjected to a 50% height reduction is examined. The initial state of the materials is characterized by electron backscatter diffraction (EBSD) data. The initial data was used to create 40 numerical simulations that were combined to arrive at a statistically comparable data set. The results suggest very good agreement between the predicted and experimental textures for both the materials within the composite.

Metallic-based composite materials have long been recognized as a way to enhance the performance of metals and continue to receive attention as a way to design new materials for novel applications. This material design process has been facilitated by the recent rapid development of nanometer-scale mechanical probes, high-resolution imaging techniques, and advanced theory and computational tools. These metallic multilayer composites have typically been produced by physical vapor deposition techniques producing thin foils, or by traditional cold-worked rolling processes with intermediate annealing steps producing plates [1,2]. The latter of these processes is achieved through an accumulated series of rolling passes and is able to produce a practical amount of plate material and therefore has the better potential for commercial manufacture of bulk materials. It is this accumulated roll bonding (ARB) process that we wish to examine here—both the resulting material and its evolution of properties. In particular, this work presents efforts to represent the evolution of the structure of the individual layers

and predict the evolution of crystallographic texture for a single rolling pass at a point in time late in the process but before the layer thickness is small enough to significantly alter the dislocation behavior within the layers.

Over the past two or three decades crystal plasticity theory and simulation capabilities have developed into a commonly

used and successful predictive tool for large plastic deformation of metallic materials [3]. Here we begin to apply crystal plasticity theory to the study of layered composite materials to examine the evolution of crystallographic texture well above the layer thickness where confined layer slip and interface dominance become important. This will

demonstrate the applicability of existing theory to problems of this type and also provide insight into the kinematics and kinetics of composite deformation through the rolling process.

Many approaches have been implemented for generating synthetic or digital representations of polycrystalline microstructures. Here we follow on the work of Sintay [4,5] and extend it to include 3D layered materials as the foundation for the 2D simulations. Initial textures were assigned to the virtual grains using data from the five 200 $\times$ 200  $\mu\text{m}$  EBSD scans (see Fig. 1) comprising a total area of  $2\times 10^5 \mu\text{m}^2$ . A total area of  $2\times 10^5 \mu\text{m}^2$  is represented by the combined 40 simulations. The dislocation state of each layer (which cannot be assumed to be fully annealed) is represented through nano-hardness measurements.

A schematic representation of the isochoric boundary conditions used for the numerical simulations is given in Fig. 2. Each of the numerical models assumed an initial aspect ratio of 1:2. Plane strain was used as an approximation of the actual rolling process and therefore six-noded plane strain triangular hybrid elements were used within commercial finite element method software (ABAQUS) [6]. A constant displacement rate was imposed on the top surface to produce a 50% height reduction in 5 seconds. The simulations were performed isothermally at a temperature of 298 K.

Crystallographic orientation at each of the material points within the 40 models were combined to give the resulting pole figures in Fig. 3. In all, these results contain sampling of 420 Cu grains and 395 Nb grains for the combined numerical simulations. In general, the numerical technique developed for this layered composite system represents the

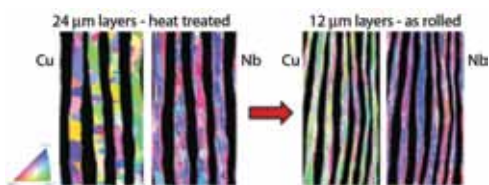


Fig. 1. EBSD data for sections of the heat-treated and rolled layered composite materials examined in this study. The rolling direction is vertical and the transverse direction is out of the page.



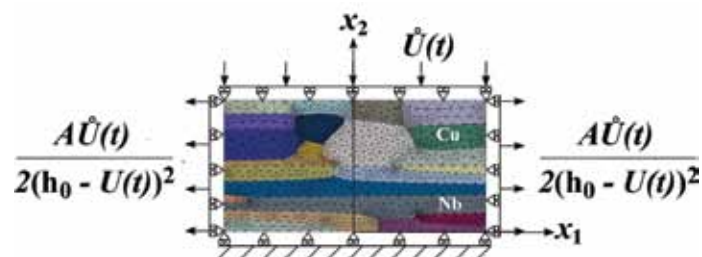


Fig. 2. Plane strain compression isochoric boundary conditions applied to a typical bilayer model with an initial 1:2 aspect ratio. The quantities  $A$  and  $h_0$  are the surface area and initial height of the numerical model.  $X_1$  corresponds to the rolling direction (RD),  $X_2$  the normal direction (ND), and  $X_3$  the transverse direction (TD).

An example of the morphology and internal stress developed as a function of height reduction is given in Fig. 4 for one realization. It is interesting to note the discrepancy between the vonMises stress developed between the two layers. This suggests that the mechanical response of the composite introduces unique boundary conditions at the interface that allow for substantially different equivalent stress states in the two different materials. The simulation results also demonstrate the substantial evolution in interface shape, which qualitatively agrees with experimental observation.

A technique for representing the processing response of bi-metallic layered composite materials was presented here and successfully used to predict the single-pass rolling texture in each layer of the Cu/Nb system reduced from 24 to 12  $\mu\text{m}$  average layer thickness. Each layer was represented by a statistically equivalent polycrystal morphology based upon EBSD data for each of the layered materials at the heat-treated 24- $\mu\text{m}$  thickness condition. Nano-indentation experiments performed on the independent material layers for the heat-treated 24- $\mu\text{m}$

experimental data well. In particular, the ability to represent the large transition in texture for the Cu but little for the Nb is noteworthy.

condition were used to initialize the single crystal model for un-annealed conditions between rolling passes. Eight different morphological realizations and, within each of those, five different crystallographic realizations, resulted in a total of 40 numerical realizations used to compare to the experimental crystallographic texture data. This number proved to be adequate to provide the proper statistical diversity to allow for representation of the initial morphology and crystallographic texture for these layered composite systems.

### Special Thanks

For important discussions with J. Wang, A. Misra, A. D. Rollett, D. L. McDowell, T. M. Pollock, T. Lookman.

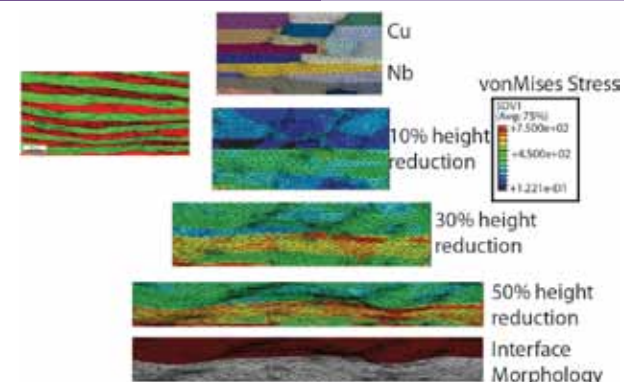


Fig. 4. Simulated deformation for one realization showing the evolution of vonMises stress with height reduction. Morphology of the experimental as-rolled 12- $\mu\text{m}$  layer thickness microstructure is shown in the phase map to the left where green is Cu and red is Nb.

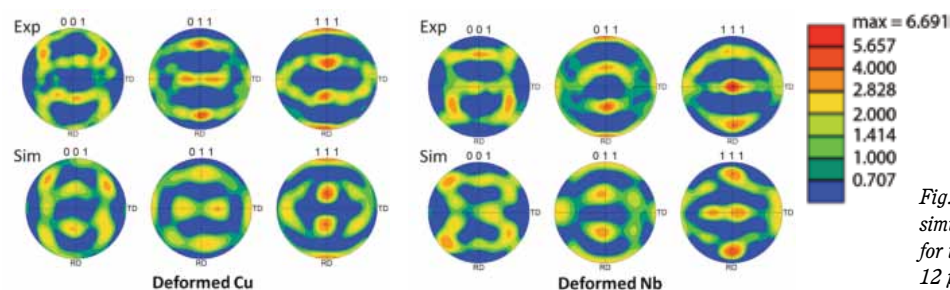


Fig. 3. Experimental and simulated equal area pole figures for the deformed Cu and Nb at 12  $\mu\text{m}$  layer thickness.

- [1] Raabe, D. et al., *Scripta Metall* **27**, 211 (1992).
- [2] Carpenter J.S. et al., *Acta Mater* **60**, 1576 (2012).
- [3] Bronkhorst, C.A. et al., *J Mech Phys Solids* **55**, 2351 (2007).
- [4] Sintay, S.D., Ph.D. "Statistical microstructure generation and 3D microstructure geometry extraction," Dissertation, Carnegie Mellon University (2010).
- [5] Sintay, S.D. et al., *3D Reconstruction of Digital Microstructures. Electron Backscatter Diffraction in Materials Science*, Springer, 139 (2009).
- [6] ABAQUS, Version 6.11 User's Manual, Dassault Systemes Simulia Corp. (2011).

# Hot Spots in the Electronic Spectrum Renormalize the Electronic Band Structure of Actinides

Tanmoy Das,  
Jian-Xin Zhu,  
Matthias J. Graf, T-4;  
Tomasz Durakiewicz,  
John J. Joyce, MPA-CMMS

Can electrons in metallic actinides be treated as non-interacting particles or must correlations and interactions between them be taken into account, and to what extent? This is among the most fundamental questions for actinide research. The answer to this question will have far-reaching consequences. For example, it will allow us to control the functionality and thermo-physical properties of nuclear fuel. Electronic correlations, while difficult to tackle, exist in a wide class of materials, including high-temperature superconductors, itinerant-electron magnets, and heavy-fermion materials. One of the common features of these strongly correlated metallic systems is the dual nature of conduction electrons. They can behave collectively as if they had a mass many times heavier than a free electron or large magnetic moment if localized. The goal for predictive materials science is to tease out common, fundamental threads that link the localized-delocalized behavior to electronic hot spots and how they affect physical properties and functionality.

**D**ynamic correlation effects of strongly correlated electron systems pose a serious challenge to our theoretical understanding of the physical behavior of actinides. Strong renormalization of electronic bands and their spectral weight anomaly cannot be accounted for by the density functional theory. The interplay between  $5f$  electrons with the conduction electrons pushes the system to the intermediate Coulomb- $U$  region in which neither purely itinerant mean-field theory nor strong-coupling Anderson lattice model hold. This twilight zone of intermediate coupling gives rise to prototypical examples of strong correlations and the emergence of multiple competing phases of matter.

The metallic U-115 compounds are often contrasted with the rich phase diagram of the isostructural Pu-115 compounds  $\text{PuCoGa}_5$ ,  $\text{PuRhGa}_5$ , and  $\text{PuCoIn}_5$ , which are superconducting with varying transition temperatures  $T_c$ . On the other hand,  $\text{UTGa}_5$  ( $T$ : Ni, Pd, Pt) compounds are antiferromagnetic, while  $\text{UCoGa}_5$  is paramagnetic. Due to this diversity of ground states in the U-115 and Pu-115 families and their suitability for experiments, they provide an excellent test bed for applying the self-consistent fluctuation approach and validation through experiment. By discovering the design principles for large spin and phonon fluctuations, we hope to correlate specific electronic “hot spots” in the spectral function with unique ground and excited state properties for controlled functionality of optical, thermal, and transport properties in actinides.

We start with ab initio electronic band structure calculations, which are then used as input to the self-consistent spin-fluctuation self-energy calculations [1,2]. This provides dressed electronic bands and quasiparticles with temperature dependence. The final calculations are validated by comparison with angle-resolved photoemission spectroscopy

(ARPES). The ARPES data reveal strong spectral weight redistribution in the single-particle spectrum with a prominent peak-dip-hump structure, with dip around 0.5 eV in Pu-115 and U-115 materials. The separation between itinerant (peak) and incoherent (hump) states is often assigned to the duality of the  $5f$  electrons. Here instead, we interpret these features in the spectrum in terms of the spin-fluctuation interaction, which creates a dip in the dressed quasiparticle spectrum due to strong scattering in the particle-hole continuum. The lost spectral weight (dip) is partially distributed to the renormalized itinerant states at the Fermi level (peak), as well as to the strongly localized incoherent states at higher energy (hump). The self-energy dressed electronic dispersions are shown in Fig. 1 for all four compounds as function of energy and momentum [1]. In all spectra the pile-up of spectral weight (hot spots) is clearly visible. At low energies all quasiparticle states are renormalized toward the Fermi level, where states are coherent and itinerant. Further away from the Fermi level, the quasiparticle states are pushed to lower binding energy. This means that the lost spectral weight is re-distributed to both lower and higher binding energies. A similar spectral weight redistribution occurs at the second peak (hump). As a result further pile-up of spectral weight occurs around 1.0-1.5 eV, creating hot spots of new quasiparticle states due to electronic correlations. Qualitatively similar behavior was also found by using the local density approximation combined with dynamical mean-field theory (LDA+DMFT) method shown in Fig. 1(e), however, with a weaker renormalization, and significantly less spectral weight near the Fermi level [3].

Our ARPES measurements in  $\text{UCoGa}_5$  validated the predicted anomalous momentum and energy dependence of the electronic dispersion [2] (see

Fig. 2). A drastic departure of the electronic states from the ab initio calculations (dashed lines) is manifest. More importantly, the associated quasiparticle width at the peak positions is significantly momentum- and energy-dependent. The energy distribution curve (EDC) of the ARPES intensity at several representative fixed momenta in Fig. 2(c), and momentum distribution curve (MDC) at several fixed energy points in Fig. 2(d) indicate that the anomaly is markedly different in energy and momentum space, which is a hallmark of electron correlations.

We conclude that the intermediate Coulomb- $U$  coupling method of spin fluctuations provides first-principles based predictive modeling capabilities of the electronic band structure in metallic actinides. Modifying optical, thermal, and transport properties in actinides by tuning correlations will be the next step toward the ultimate goal of controlled functionality in actinide materials.

Fig. 1. (a)-(d): The self-energy renormalized angle-resolved spectral weight along high-symmetry directions in momentum space. The peak-dip-hump feature is evident in all spectra between 0 to 1 eV. Two arrows in panel (b) indicate generic dips in the spectral weight, related to peaks in the self-energy. For comparison, the hot spots in the dynamical mean-field theory (DMFT) calculations of  $\text{PuCoGa}_5$  by ref [3] are shown in (e).

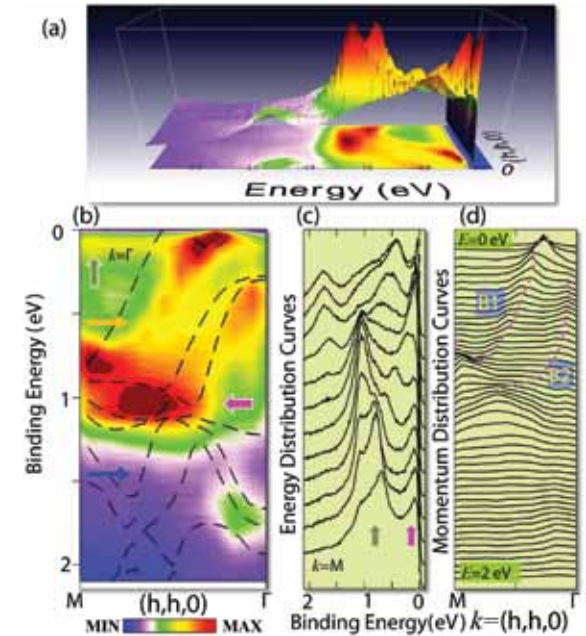
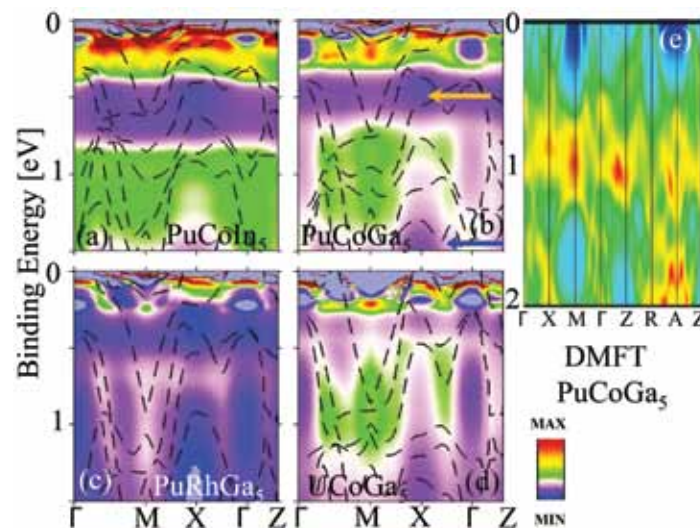


Fig. 2. Measured ARPES spectral function anomalies. (a) 3D intensity map of  $\text{UCoGa}_5$  along the  $M \rightarrow \Gamma$  direction  $(h, h, 0)$  in the Brillouin zone. (b) Same data plotted as 2D contour map and compared to the corresponding ab-initio electronic band structure dispersions (black dashed lines). Arrows in (b) indicate two quasi-non-dispersive energy scales of higher intensity as well as kink. (c) The EDC is plotted for fixed momenta. The curves from bottom to top are chosen for equally spaced momenta from  $M$  to  $\Gamma$ . Arrows have the same meaning as in panel (b). (d) The MDC for fixed binding energy. Bottom to top curves are chosen from binding energies  $E = 0$  to  $2$  eV. The dashed lines are guide to the eyes for two extracted low-lying dispersions of anomalous character.

- [1] Das, J. Zhu, X., Graf, M.J. *Phys Rev Lett* **108**, 017001 (2012).
- [2] Das, T. et al., *Phys Rev X* **2**, 041012 (2012).
- [3] Pourovskii, L.V. et al., *Phys Rev B* **73**, 060506(R) (2006).



# An Efficient Kernel Polynomial Method for Calculating Transition Rates in Large-Scale Materials

Chen Huang,  
Arthur F. Voter,  
Danny Perez, T-1

We developed an efficient method for calculating transition rates in large-scale materials. In this method, we first reformulated the prefactor of the transition rate in terms of the density of states (DOS) of the Hessian matrix and then solved for the DOS using the kernel polynomial method. We demonstrated this novel method by calculating the prefactor of the vacancy hop in bulk silver and compared the results to the benchmark that was obtained by directly diagonalizing the Hessian matrix.

The transition rates of important kinetic processes in materials are essential for our understanding and prediction of material properties. For example, the dislocation nucleation rate controls plasticity, which determines how much strain materials can sustain. A recent discovery [1] of a self-healing mechanism shows that grain boundaries after radiation can emit interstitial defects to annihilate the vacancy defects nearby. Such a self-healing mechanism might be relevant for the design of next-generation nuclear reactors. It is then important to know how fast such a self-healing process operates, especially when it is also competing with other processes. However, it is usually too computationally expensive to calculate transition rates in large-scale materials. Within harmonic transition state theory [2], one must calculate all the eigenvalues of the Hessian matrices at the saddle point and the basin minimum. The computational cost then scales cubically with respect to system size. Unfortunately, even a simple atomistic simulation of dislocation nucleation requires thousands of atoms, which in turn produces very large Hessian matrices that can be difficult to diagonalize.

In this work, we present a novel method for calculating transition rates in large-scale materials. To avoid diagonalizing the Hessian matrix, we reformulate the transition rate in terms of the DOS of the Hessian. To efficiently solve for the DOS, we employ the kernel polynomial method (KPM) [3] in which the DOS is expanded with Chebyshev polynomials. The expansion coefficients, that is, the moments, are then obtained with stochastic sampling. In the past, KPM has been successfully applied to the calculation of the DOS in quantum mechanics simulations [4]. To our knowledge, this is the first time the KPM is used in the calculation of transition rates in large-scale materials. Assuming that the required number of moments is constant and assuming short-ranged potentials, our method is linear-scaling with respect to problem size, which makes

it very promising for application to large-scale systems. Our method also offers a good balance between accuracy and computational cost.

We demonstrate our method by computing the rate prefactor for a simple vacancy hop in bulk silver. Tests on more complicated processes are underway. In Fig. 1, we show the saddle point configuration. For simplicity, we show only the atoms around the vacancy, with all bulk atoms removed. Atom A is moving to the vacancy B and creating a new vacancy C. We calculate the prefactor of this transition with KPM and compare it against the benchmark obtained by directly diagonalizing Hessian matrices. In KPM, the zero-th moments are simply equal to the dimensions of the problems; we therefore can fix the zero-th moments to their exact values by properly constructing the random vectors used in the calculation of the moments. In Fig. 2, we show that the convergence of the prefactor with respect to the number of random vectors using exact zero-th moments (red crosses) is much faster than the case without using exact zero-th moments (green circles). In a recent work [5], the prefactor was calculated using thermodynamic integration in which molecular dynamics is performed on model harmonic potentials associated with the saddle point and the basin minimum. Compared to that work, one prominent advantage of our approach is that we can achieve significant error cancellation between the partition functions at the saddle point and the basin minimum by using the same sequence of random vectors for each. In Fig. 2, the blue triangles are calculated using two different sequences of random vectors and converge much slower than the case with matched random vectors. In the lower plot of Fig. 2, we show an even slower convergence using inexact zero-th moments and non-matched random vectors. In Fig. 3, we show the convergence speed of the prefactor with respect to the number of moments. The prefactors converge to better than 10% error with less than ten moments. Such fast convergence makes our method very appealing. In Fig. 4, we show the convergence of prefactors with respect to random vectors for five

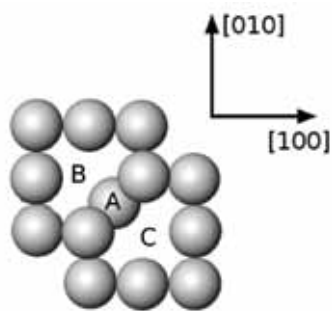


Fig. 1. Saddle point configuration for vacancy hopping. Atom A is moving to vacancy B and creating a new vacancy C. For simplicity, we performed the common neighbor analysis and removed all atoms in face-centered cubic environment.

different cell sizes. The 10% errors are drawn with dashed lines. For all cell sizes, the prefactors converge to the  $\pm 10\%$  window with fewer than 3,000 random vectors.

In summary, we have developed a powerful method for calculating transition rates of important processes in large-scale materials. Great insight should be accessible with this method, for processes such as the migration speed of dislocations in metals under shock-waves, which controls the flow of plasticity, the unexpected short lifetime of dislocations in metal nanocrystals, which is responsible for their ultra-strength, and the aggregation of small voids to form large voids under stress, which causes the failure of materials.

Fig. 2. Comparison of the convergence speeds of prefactors (using 400 moments) with respect to the number of random vectors for four different schemes. (Upper plot) We show results from (1) exact zero-th moments and matched random vectors (red crosses), (2) exact zero-th moments and non-matched random vectors (blue triangles), and (3) inexact zero-th moments and matched random vectors (green circles). (Lower plot) Results from inexact zero-th moments and non-matched random vectors. The fluctuation is much larger than the other cases. Exact prefactors are shown by dashed lines in both plots.

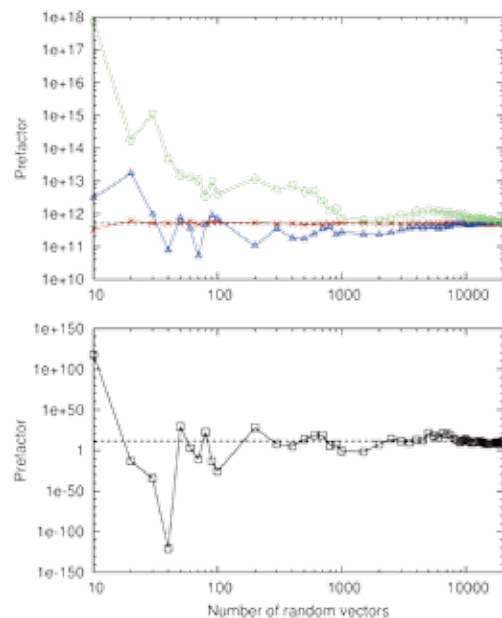


Fig. 3. Convergence of prefactors (using 20,000 random vectors) with respect to the number of moments in the Chebyshev expansion, for five different cell sizes. We find a very fast convergence of prefactors by increasing the number of moments. With 10, 20, 30, 40, and 50 moments, prefactors converge to 8%, 4%, 2%, 0.8%, and 0.6%, respectively.

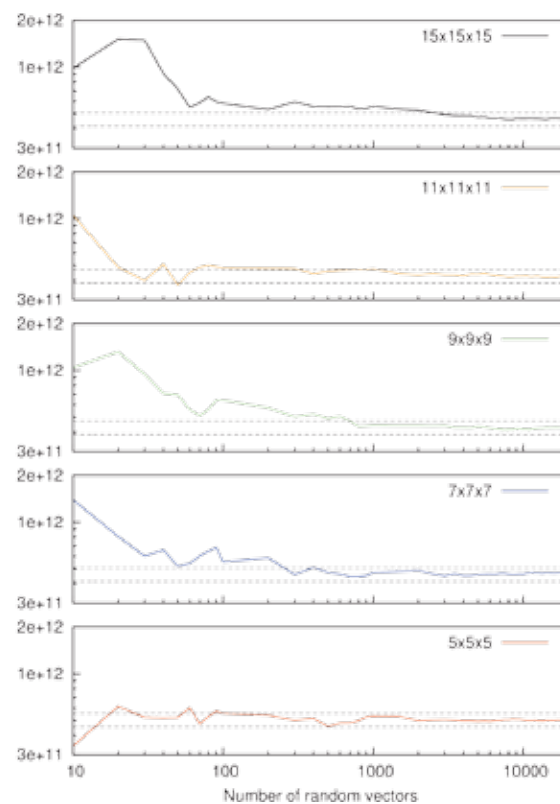
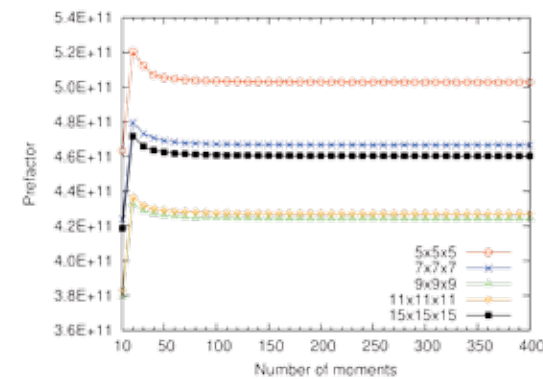


Fig. 4. Convergence of prefactors (using 400 moments) with respect to the number of random vectors, for five different cell sizes. The dashed horizontal lines show the  $\pm 10\%$  errors from the benchmark which is calculated by directly diagonalizing the Hessian matrices (except for the 15x15x15 cell, where the benchmark is taken from the converged KPM result at 20,000 random vectors).

- [1] Bai, X.-M. et al., *Science* **327**, 1631 (2010).
- [2] Vineyard, G.H., *J Phys Chem Solids* **3**, 121 (1957).
- [3] Weiße, A. et al., *Rev Mod Phys* **78**, 275 (2006).
- [4] Voter, A.F. et al., *Phys Rev B* **53**, 12733 (1996).
- [5] Jin, C. et al., *Scripta Mater* **62**, 206 (2010).

# Atomic Interface Design of Nanocomposites with Superior Shock Tolerance

Ruifeng Zhang, T-3;  
Timothy C. Germann, T-1;  
Jian Wang, MST-8;  
Irene J. Beyerlein, T-3

A new empirical potential for a Cu/Nb system is developed that accurately captures deformation twinning in body-centered cubic (bcc) metals and the mechanical response of bimaterial incoherent Cu/Nb interfaces in the extreme conditions of shock compression. Using this potential, we demonstrate the important role of atomic interface structure on the dislocation-mediated plastic response of Cu/Nb nanolayered composites in shock compression. The activation barriers for the nucleation of dislocations from and the transmission of dislocations across an atomically flat interface are found to be substantially higher than those from an atomically faceted interface (that is, an interface comprised of a regular array of steps).

For next generation advanced materials, the scientific and technical demand has been increasing for a design such that the next generation advanced materials that can withstand extreme mechanical loadings, such as severe plastic deformation (SPD) and shock loading. Promising candidates are nanostructured composite materials with a high bimaterial interfacial content. In such nanomaterials, the atomic interface structure deviates significantly structurally and/or chemically from that of the bulk crystals that they join. Under extreme mechanical loadings, the interfaces can serve as a barrier to plastic flow, act as a source of dislocations, and accommodate deformation by rotation or sliding, etc. Different interfacial defect characteristics can alter the way the interface responds to extreme mechanical loadings.

To elucidate this unique structure-property relationship, a realistic atomistic simulation plays an important role in revealing how the interface responds to the mechanical loadings and interactions with dislocations. In our recent studies, we have focused our modeling efforts on the face-centered-cubic (fcc)-bcc nanocomposite (e.g., Cu/Nb nanolaminate composites) as a prototype system. We found that an obstacle for achieving realistic atomistic simulations in shock loading for bcc metals, such as Nb, and fcc-bcc binary systems, such as Cu-Nb, was the inaccuracy of existing potentials in the high-pressure regime. Current potentials lead to an artificial structural phase transformation under high pressure for bcc-Nb that disagrees with density functional theory (DFT) calculations and experiment observations. Thus, they are inappropriate for shock compression studies (Fig. 1) [1]. We discovered that the source of the problem was the existence of artificial minima in the high-pressure gamma surface. At 50 GPa, only the Ackland Finnis-Sinclair (FS) potential and force-matched embedded atom method (EAM) potentials can successfully capture the associated changes in the gamma surface

predicted by DFT. Such findings set a foundation for reliable atomistic shock simulations that involve materials that are bcc or contain a bcc constituents, such as two-phase Cu-Nb composites. Using the elemental Ackland potential for Nb and Voter-Chen Cu potential, we developed a new Cu/Nb interatomic cross potential, which can accurately described both Cu/Nb interface structure and the high-pressure response of the Cu and Nb constituents.

Utilizing the newly constructed Cu/Nb potentials and non-equilibrium molecular dynamics (MD) simulations, we successfully revealed the dislocation processes underlying the pronounced effect of atomic interface structure on the plastic response of Cu/Nb nanolayered composites to shock compression [2,3], which was recently observed in experiments. Furthermore, we demonstrated that the critical shock pressures to nucleate and transmit dislocations across an atomically flat interface (e.g., KS interface) is substantially higher than those from a faceted interface (e.g., {112}KS interface) [2]. The reasons can be related to atomic-level interface dislocation characteristics that cause these two interfaces to nucleate, absorb, and transmit dislocations by significantly different mechanisms (Fig. 2). We have also studied the effect of layer thickness on the mechanisms of shock deformation. The study reveals an inverse size-dependence. We find enhanced shock resistance with decreasing layer thickness above a critical layer thickness, while below this thickness the shock resistance deteriorates with decreasing layer thickness due to the onset of large amounts of dislocation nucleation from the interfaces. These results will be presented in a forthcoming paper.



Fig. 1. Two types of structural changes that develop in an Nb single crystal under shock loading: (a) deformation twinning using the A.MFS potential and (b) bcc-fct phase transformation using the D.EAM potential, (c) a magnification of the twin structure, and (d) a magnification of the fct structure. The GSF curves on the {112} plane along the  $\langle 111 \rangle$  direction at (e) zero hydrostatic pressure and (f) 50 GPa hydrostatic for the five empirical potentials and DFT. Local minima are found for the J.EAM, D.EAM, D.EFS potentials.

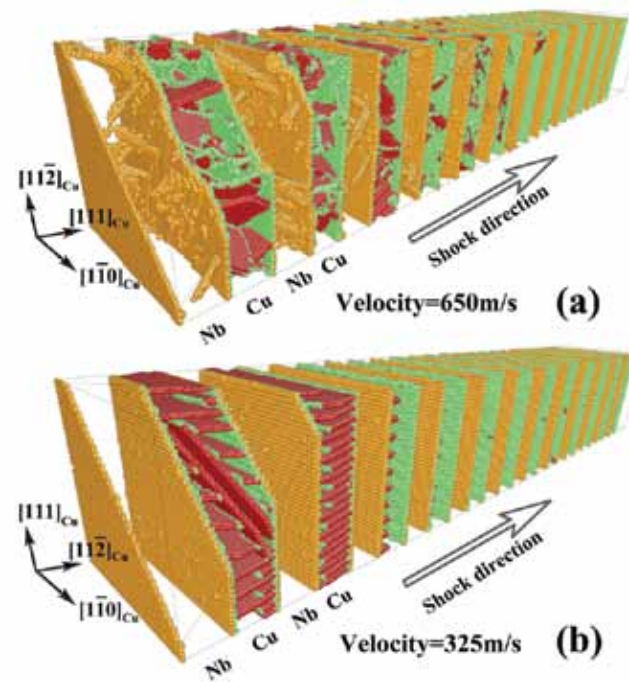
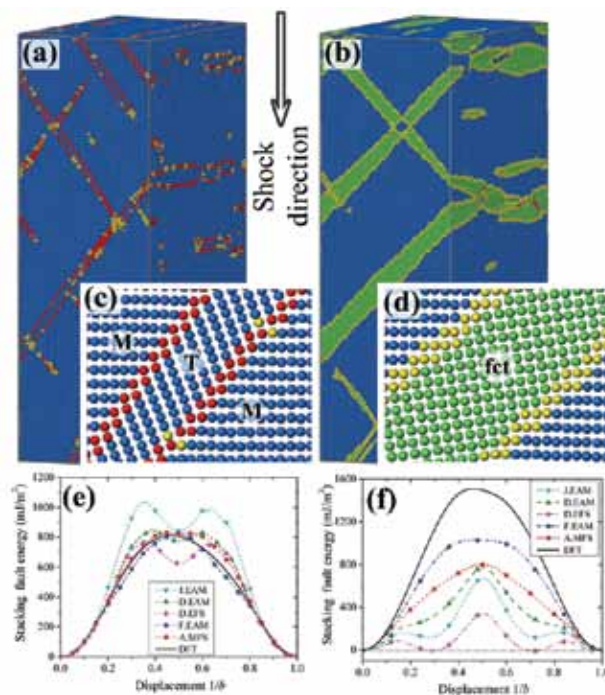


Fig. 2. Perspective view of shock-induced dislocation nucleation and transmission [2,3] from (a) flat KS interfaces at the piston speed of 650 m/s and (b) faceted {112}KS interfaces at  $u_p = 325$  m/s during shock compression.

[1] Zhang R.F. et al., *Phil Mag Lett* **91**, 731 (2011).

[2] Zhang R.F. et al., *Scripta Mater*, in press; doi: <http://dx.doi.org/10.1016/j.scriptamat.2012.09.022> (2012).

[3] Zhang R.F. et al., *AIP Conf Proc* **1426**, 1251 (2012).

# Development of Interface-Dislocation Dynamics Simulations for Nanoscale Metallic Composites

Caizhi Zhou, CNLS/T-3;  
Jian Wang, MST-8;  
Irene J. Beyerlein, T-3

Extensive investigations over the past decade indicate that nanoscale metallic composites have unprecedented levels of strength, ductility, and damage tolerance in extreme environments. However, the current meso-/macro-scale materials modeling cannot predict the mechanical response of such nanoscale composites because they do not account for the role of the interface mechanical deformation. In this work, we apply findings from atomistic simulations to develop an Interface-Dislocation Dynamics (DD) simulation code with the focus on capturing the fundamental physics of dislocation-interface interactions, such as annihilation, nucleation, and emission within, at, and across the interface. With this model, we can study the influence of interface structure and properties and the misorientation on the development of plastic deformation in nanoscale composites. These findings will enable the development of material modeling tools at continuum-length scales.

**M**ultilayered materials, typically synthesized by physical vapor deposition or electro-deposition techniques, exhibit significant increases in strength as the layer thickness is decreased from the micrometer to nanometer scale [1]. It has been widely recognized that materials properties can be significantly improved or degraded by virtue of interface structures [2]. Atomic-scale modeling is able to reveal the unit process (involving single or a few defects) during deformation with respect to kinetics and energetics, but is limited to timescale (ns) and lengthscale (nm) [3]. Current state-of-the-art continuum modeling can only treat interface boundaries phenomenologically without any structural characteristics [4]. For materials containing high-density interfaces, it is still a challenge in incorporating interface physics, such as nucleation, motion, reactions, etc., in current models. The bottleneck is attributed to three reasons: (1) lack of knowledge of interface-dominated deformation mechanisms, in particular of dislocation-interface interactions; (2) slips in the adjacent grains are not directly correlated to the interface with respect to their orientation relations, interface structures, and interface properties; and (3) a generic interface model representing the structure and properties of interfaces does not exist.

DD simulations, in which dislocations are the simulated entities, offer a way to extend length and time scales beyond those of atomistic simulations, and have the unique advantage of exploring dislocation activities [5] due to (1) the powerful capability of modeling the behavior of metallic materials at the micro-scale in a more physical manner than existing continuum plasticity models, and (2) direct observation of dislocation activities in simulated samples. In most studies, DD simulations have been implemented on single-crystal materials, such as investigations of strain hardening of bulk materials under low and high strain rates [6], dislocation motions and interactions in thin films [7],

and size-dependent plasticity of single-crystal nano-/micro-pillars [8]. In this work, by combining atomistic studies with DD simulation, we developed an Interface-DD model for metal-metal interfaces to overcome the weakest areas in the DD simulations—in accounting for the special roles of interfaces on storage, recovery, nucleation, and emission of dislocations within, at, and across interfaces. The development of the interface-DD model relies heavily on the understanding of dislocation-interface interaction at atomic scale.

In 3D discrete DD simulations, dislocations are discretized into segments, as shown in Fig. 1. Each segment contains the information of dislocations, such as Burgers vector, line direction, forces, and mobility. The total force on each segment is calculated by the Peach–Koehler (PK) equation:

$$\mathbf{F}_{PK} = \left( (\boldsymbol{\alpha}_{app} + \boldsymbol{\sigma}_{int}) \cdot \mathbf{b} \right) \times \mathbf{t} + \mathbf{F}_{self} \quad (1)$$

where  $\boldsymbol{\sigma}_{app}$  is the applied stress tensor,  $\boldsymbol{\sigma}_{int}$  is the stress tensor from the other defects,  $\mathbf{b}$  is the Burgers vector of the dislocation,  $\mathbf{t}$  is the line direction of a given segment, and  $\mathbf{F}_{self}$  is the self-force of the dislocation segment.

During the deformation, the plastic strain rate is obtained from the motion of dislocations as

$$\dot{\epsilon}^p = \frac{1}{2V} \sum_{i=1}^{N_{tot}} l_i^\alpha v_i^\alpha (\mathbf{b}_i \otimes \mathbf{n}^\alpha + \mathbf{n}^\alpha \otimes \mathbf{b}_i) \quad (2)$$

where  $V$  is the volume of the simulated crystal,  $N_{tot}$  is the total number of dislocation segments,  $l_i^\alpha$  is the length of dislocation segment  $i$  moving on the slip plane  $\alpha$ , and  $v_i^\alpha$  is the corresponding moving velocity of the segment  $i$ .  $\mathbf{b}_i$  and  $\mathbf{n}^\alpha$  are the Burgers vector of dislocation segment  $i$  and the normal of slip plane  $\alpha$ , respectively.

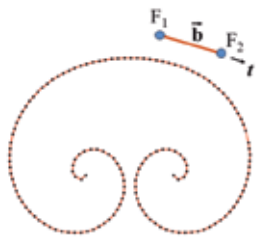


Fig.1. Illustration of the discretization of a Frank-Read dislocation source in 3D DD simulations.

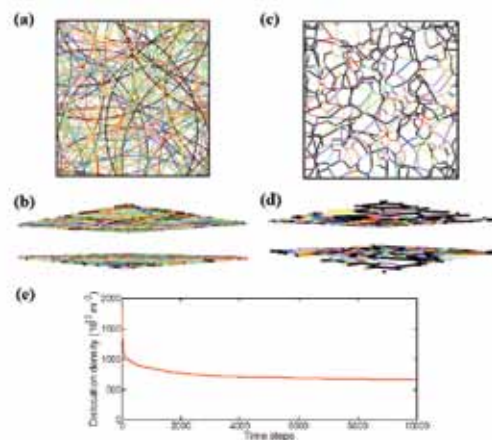


Fig. 2. (a) Top view of initial dislocation structure for bilayer Cu/Nb system, (b) side view of (a), (c) top view of relaxed dislocation structure for bilayer Cu/Nb system, (d) side view of (c), and (e) the evolution of dislocation density during relaxation.

Atomistic simulations so far reveal that (1) interface sliding can be represented as a result of the nucleation and glide of glide-type interfacial dislocations, (2) lattice dislocations can nucleate from interfaces [9], (3) all types of interfacial dislocations can move through either glide or climb, and (4) once lattice dislocation enters interfaces, it becomes an interfacial dislocation (conserves the Burgers vector but loses the slip plane). As a consequence, it moves along the interface through either climb or glide, and acts as sources for nucleating lattice dislocation. In keeping with these findings, we develop the rules of the nucleation and reaction of interfacial and lattice dislocations in the interface-DD model.

With the dislocation-interface interaction rules developed in our DD-interface model, we studied the mechanical response of a Cu/Nb bilayer system under external load. Initially, dislocation loops were generated in both and the Burgers vectors for these loops were randomly chosen from Burgers vectors in both Cu and Nb phases as shown in Fig. 2(a-b). Then the system was relaxed to equilibrate the dislocation microstructure. Figures 2(b-c) and (d) show the final equilibrated dislocation structure and the evolution of the dislocation density during relaxation, respectively. After relaxation, an external load much like that in the experiment was applied on the system. Driven by the external load, lattice dislocations nucleated from the interface to accommodate the applied deformation. After nucleation, lattice dislocations continued propagating in phases due to the operation of Frank-Read sources. Once approaching another interface, the front part of the half dislocation loop was absorbed by the interface and two threading dislocations were formed after the front part of the dislocation loop was truncated by the interface. Under

external load, these two threading dislocations propagated in the Cu phase and generated plastic strain to minimize the internal strain energy. Figure 3 (a) shows one example of the whole process of the nucleation of lattice dislocations from interfaces and the propagation of threading dislocations in the Cu phase. With increasing load, lattice dislocations continued nucleating from the interface and generated large amounts of plastic deformation in the system. Figure 3 (b) shows the dislocation structure of the deformed system at 2% total strain.

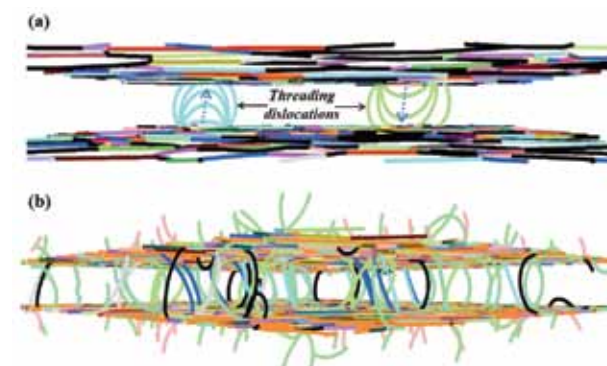


Fig. 3. (a) Progress of lattice dislocations nucleating from interfaces and the formation of threading dislocations after nucleated dislocations absorbed by the interface (dashed arrows indicate dislocation operating direction), (b) plot of dislocation structure after deformation at 2% total strain.

In this work, we developed a DD simulation model based on observations from atomic-scale simulations and experimented to study the dislocation-interface interactions at micro- and nano-scales, and provide a generic bi-metal interface model that represents the key characteristics of the structure and properties of the interface. This model can also be applied to dislocation interactions with grain boundaries and twin boundaries, since the basic physical processes such as dislocation absorption, nucleation and transmission from interfaces are the same. Our improvement on current DD simulations removes the restriction on single-phase materials and fills the gap in materials modeling on interface problems between atomic-scale models and continuum plasticity models.

- [1] Misra, A. et al, *Acta Mater* **53**, 4817 (2005).
- [2] Demkowicz, M.J. et al., *Dislocat Solid* **14**, 141 (2008).
- [3] Wang, J. et al., *Acta Mater* **56**, 5685 (2008).
- [4] Bronkhorst, C.A. et al., *J Mech Phys Solid* **55**, 2351 (2007).
- [5] Ghoniem, N.M. et al., *Phys Rev B* **60**, 128 (1999).
- [6] Wang, Z.Q. et al., *Int J Plast* **25**, 26 (2009).
- [7] Pant, P. et al., *Acta Mater* **51**, 3243 (2003).
- [8] Zhou, C. et al., *Acta Mater* **58**, 1565 (2010).
- [9] Zhang, R.F., et al., *Scripta Mater* **65**, 1022 (2011).



# Jamming, Pattern Formation, and Dynamic Phases for Driven Dislocation Assemblies

Caizhi Zhou, CNLS/T-3;  
Irene J. Beyerlein, T-3;  
Charles Reichhardt, T-4;  
Cynthia Olson-Reichhardt, T-1

Dislocation structures within individual crystals organize into patterns ranging from tangles to cells to planar walls, becoming more refined as stress or strain increases. Under a drive, pileups and intermittent dynamics arise near the depinning transition, but correlations between patterning and the intensity of applied stress or strain have not been established. We demonstrate that driven dislocation assemblies exhibit the same non-equilibrium phases as those observed for collectively interacting particle systems, such as vortices in superconductors or sliding charge-density waves. The analogous phases are a jammed state below yielding, a strongly fluctuating intermediate state above yielding, and a quasi-ordered phase at higher drives, detectable via dislocation structure, mobility, velocity distribution, and velocity noise. This implies that many established results obtained for driven vortices can be applied to dislocation dynamics.

The dynamics and pattern formation of dislocations are of tremendous importance for understanding materials properties. Driven dislocations are also an outstanding example of a non-equilibrium many-body system where a number of competing interactions come into play. Of particular interest is how the dislocations can organize themselves into patterns under different applied loads, how these patterns can be characterized, and whether there are distinct dynamic phases as a function of load. It is known that organized dislocation structures within individual crystals, such as tangles, cells, or planar walls, can become more refined and better defined as stress or strain increases.

2D and 3D dislocation dynamics simulations based on linear elasticity theory predict self-organization of dislocation assemblies into varying configurations, such as pileups near the yielding or depinning transition [1] and 2D mobile walls [2] or 3D slip bands [3,4] under an external drive. Below a critical stress where dislocations show no net motion,

the system is considered jammed [5], while intermittent or strongly fluctuating behavior with highly jerky or avalanche-like motion occurs above the critical stress [1,4]. Avalanche behavior with power-law velocity distributions is proposed to be a signature of critical dynamics [1,4-6]. No correlations between the transitions in patterning and the intensity of applied stress or strain have been established before now. Another important problem is whether the patterning could determine how the yielding changes as a function of dislocation density.

We utilize a discrete dislocation dynamics model with periodic boundary conditions for a 2D cross-section of a sample containing straight-edge dislocations that glide along parallel slip planes. The dislocations

interact via a long-range anisotropic stress field that is attractive between two oppositely signed dislocations and repulsive for liked-signed pairs. Under an external applied stress  $\tau_{ext}$ , dislocation  $i$  moves along  $x$  in its assigned plane according to an over-damped equation of motion given by

$$\eta \frac{dx_i}{dt} = b_i \sum_{j \neq i}^N \tau_{int}(r_j - r_i) - \tau_{ext} \quad (1)$$

where  $x_i$  is the  $x$  coordinate of  $i$ th dislocation at point  $r_i = (x_i, y_i)$  with Burgers vector  $b_i$ ,  $\eta$  is the effective friction and  $\tau_{int}(r_j - r_i)$  is the long-range shear stress on dislocation  $i$  generated by dislocation  $j$ . For  $r = (x, y) = (x_i, y_i) - (x_j, y_j)$ ,  $\tau_{int}(r_j - r_i)$  for an edge dislocation with Burgers vector  $b$  is

$$\tau_{int}(r) = b\mu \frac{x(x^2 - y^2)}{2\pi(1 - \nu)(x^2 + y^2)^2} \quad (2)$$

where  $\mu$  is the shear modulus and  $\nu$  is the Poisson's ratio.

After randomly placing the dislocations in the simulation volume, the system is allowed to relax without external load, so that the dislocations can achieve an equilibrium configuration to minimize the system energy (Fig. 1(a)). With external load  $F_d$  increasing, the dislocation structure changes (Fig. 1(a)-(d)) and produces signatures in  $\langle |v| \rangle$  versus  $F_d$  as shown in Fig. 2.

Below yielding, the dislocation pattern slowly changes and gradually forms a dipolar wall (Fig. 1(b)) after each load increment but  $\langle |v| \rangle$  goes to zero in the long-time limit, indicating that the system is in the jammed phase below the critical yield. Just above yielding, the dipolar wall structure breaks down as shown in Fig. 1(c) and the system enters a state characterized by strong fluctuations in the dislocation positions.

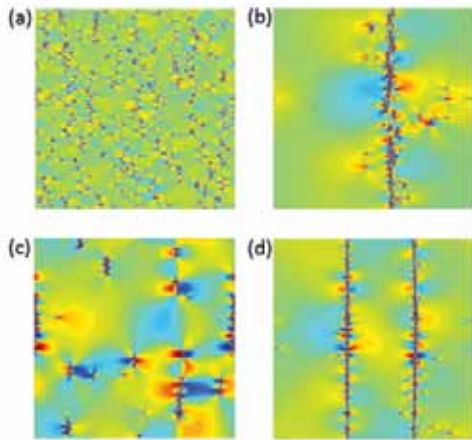


Fig. 1. Stress maps of the sample range from large negative (blue) to large positive (red) stress. (a) The initial dislocation positions at zero load. (b) Just before yielding, the dislocations are predominantly located at pile-ups to form a single bipolar wall. (c) Above yielding at  $F_d = 3.6$ , the wall breaks apart and the structure exhibits intermittent dynamics. (d) At  $F_d = 8.0$  there is a dynamical ordering into polarized walls, each composed of dislocations with the same Burgers vector orientation.

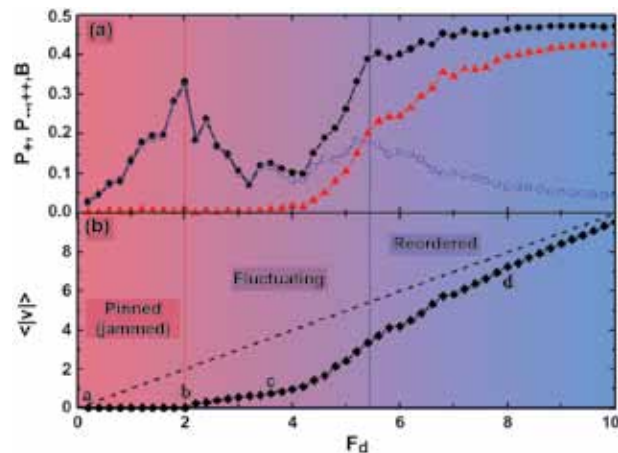


Fig. 2. (a)  $P_+$  (blue squares), the fraction of dipolar walls, versus  $F_d$  has a peak just below yielding.  $P_-$  (black circles), the fraction of uni-polar walls, passes through a plateau when the polarized wall state forms.  $B$  (red triangles) is a measure of the net Burgers vector in the walls. (b) The average absolute value of the dislocation velocity  $\langle v \rangle$  (solid lower curve) versus  $F_d$ . The upper dashed curve shows  $\langle v \rangle$  for non-interacting dislocations. Visible in the lower curve is a yielding point, a nonlinear region corresponding to the disordered or fluctuating regime, and a linear region at high drives when the system is dynamically ordered. Points a, b, c, and d indicate the  $F_d$  values illustrated in Fig. 1.

Under high drive, the dislocations form unipolar walls composed of only one type of dislocation, either negative or positive as shown in Fig. 1(d).

By conducting a series of simulations for varied dislocation densities,  $\rho$ , and analyzing the ordering dynamics, we construct the dynamic phase diagram shown in Fig. 3. The lower curve indicates the yielding transition from the low drive

jammed or pinned phase of dipolar walls to the fluctuating disordered phase. The onset of the dynamically ordered phase is defined as the force at which the unipolar wall structures start to form, and is plotted in the upper curve. As  $\rho$  increases, the yielding point rises to higher  $F_d$  since the dislocations have a more difficult time breaking through the dipolar walls that form. The increase in yield threshold with increasing  $\rho$  remains robust when we perform simulations with different initial dislocation configurations. In addition, the onset of the high-drive dynamically ordered phase also increases in a similar fashion with increasing  $\rho$ . This phase diagram exhibits the same features observed for vortex systems as a function of pinning strength versus external drive, where both the critical depinning force and the onset of the ordering rise to higher drives with increasing pinning strength [7].

To summarize, we have shown that driven dislocation assemblies exhibit several distinct non-equilibrium phases as a function of drive that are associated with the formation of distinct dislocation patterns. The jammed to yielding transition is correlated with the formation of 1D dislocation pile ups, and above yielding the system transitions into a fluctuating intermittent phase where the dislocation

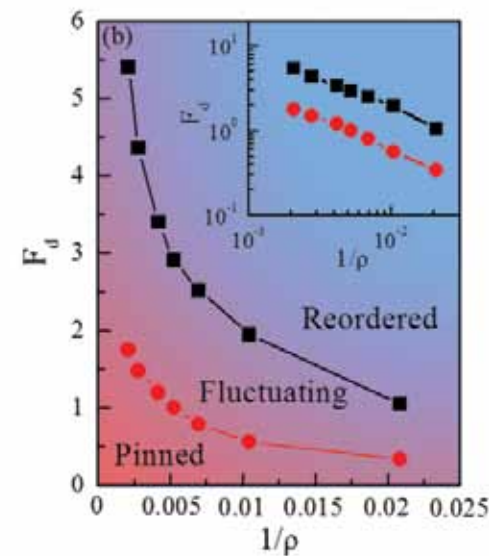


Fig. 3. The dynamical phase diagram  $F_d$  versus  $1/\rho$ , where  $\rho$  is the dislocation density. The lower curve (red circles) indicates the onset of yielding and the upper curve (black squares) is the onset of the dynamically induced ordered phase; the fluctuating phase falls between the two curves. Both the critical yielding and the dynamical ordering shift to higher drives as the  $\rho$  increases. Inset: The same curves plotted on a log-log scale.

structures break apart and reform. This is followed at higher loads by a more ordered state of moving polarized dislocation walls. All of the states are associated with transport signatures such as changes in the transport noise fluctuations as well as features in the dislocation velocity versus applied shear, in analogy with velocity-force curves. We also find that the transition from the jammed state to the polarized moving wall state moves linearly to higher load with the dislocation density. Finally, we note that many of the features described here are remarkably similar to the dynamic phases observed in driven many-body systems with quenched disorder, such as vortices in type-II superconductors.

- [1] Miguel, M.-C. et al., *Nature* **410**, 667 (2001).
- [2] Laurson, L. et al., *Phys Rev Lett* **105**, 015501 (2010).
- [3] Wang, Z.Q. et al., *Phil Mag* **88**, 1321 (2008).
- [4] Csikor, F.F. et al., *Science* **318**, 251 (2007).
- [5] Groma, I. et al., *Phys Rev Lett* **108**, 269601 (2012).
- [6] Dimiduk, D.M. et al., *Science* **312**, 1188 (2006).
- [7] Moon, K. et al., *Phys Rev Lett* **7**, 2778 (1996).

## Author Cross Reference

### A

Abdallah, Joseph Jr. ....	102
Agarwal, Bhaskar .....	30
Aleksandrov, Lyudmil .....	78
Alexandrov, Boian S.....	112
Alexandrov, Ludmil B. ....	112
Anderson-Cook, Christine .....	62
Andonov, Rumen.....	90
Andrews, Malcolm J.....	16
Armstrong, Gregory.....	102

### B

Bakosi, Jozsef .....	4
Batista, Enrique.....	124
Bauer, Amy L. ....	112
Bayless, Amanda J.....	24
Bent, John .....	82
Bent, Russell .....	92
Berendzen, Joel.....	116
Bergen, Ben .....	32
Berman, Gennady P.....	64, 98
Berndt, Markus.....	44
Beyerlein, Irene J. ....	132, 138, 140, 142
Bhat, Sham .....	38
Bishop, Alan R.....	64, 112
Bitz, Cecilia M .....	54

Blanchard, Sean.....	76
Bronkhorst, Curt A. ....	132
Burr, Thomas L.....	66
Bussod, Mira.....	116

### C

Carlson, Joseph A. ....	104
Carpenter, John S. ....	132
Carrington, David B. ....	6, 8
Cawkwell, Marc .....	128
Chapuis, Guillaume.....	90
Chen, Qingshan.....	36
Chen, Hsing-Bung.....	68
Christon, Mark A. ....	4
Chen, Shuh-Rong.....	132
Cirigliano, Vincenzo .....	100
Clements, Bradford E. ....	130
Cohn, Judith .....	116
Colgan, James P. ....	102
Coon, Ethan .....	44
Cranor, Chuck .....	82

### D

Dalla Vecchia, Claudio .....	30
Das, Tanmoy .....	134
Davis, Kei .....	70, 72, 74
DeBardeleben, Nathan.....	76



Del Valle, Sara Y. ....	86
Denissen, Nicholas A. ....	16
Ding, Xiaoning ....	72
Djidjev, Hristo ....	78, 90
Durakiewicz, Tomasz ....	134

## E

Eidenbenz, Stephan ....	94
Ellinger, Carola ....	26
Espy, Michelle A. ....	98
Even, Wesley ....	24

## F

Ferenbaugh, Charles R. ....	80
Fenimore, Paul W. ....	114
Fields, Parks ....	68
Fontes, Christopher J. ....	102
Frey, Lucille H. ....	24
Fryer, Chris L. ....	24, 26
Fung, Jimmy ....	16

## G

Gandolfi, Stefano ....	104
Garimella, Rao ....	44
Gattiker, James R. ....	38
Germann, Timothy C. ....	88, 138
Grabowski, Paul E. ....	106
Graf, Matthias J. ....	134

Gray, George T, III ....	132
Graziani, Frank R. ....	106
Grider, Gary ....	82
Goff, George S. ....	124
Gorshokov, Vyacheslav N. ....	98
Gyrya, Vitaliy ....	10
Gunzburger, Max ....	58

## H

Hackel, Peter ....	102
Hamada, Michael S. ....	66
Hamann, Bernd ....	50
Hansen, Benjamin L. ....	132
Hayes, Ryan ....	122
He, Jun ....	82
Hebert, David ....	40
Hecht, Matthew W. ....	38, 50
Heinrich, Juan C. ....	6, 8
Hengartner, Nick ....	114, 116
Hennelly, Scott ....	122
Hickman, Kyle S. ....	86
Hickmott, Donald D. ....	120
Hlavacek, William S. ....	112
Holz, Daniel E. ....	30
Hong-Geller, Elizabeth ....	112
Huang, Chen ....	136

Hunke, Elizabeth C. ....	40, 54
Hyman, James M. ....	86

## I

Ionita, Axinte ....	130
---------------------	-----

## J

Jacobsen, Doug ....	52
Jian, Stevenson ....	76
Jiang, Song ....	70, 72, 74
Johnson, Jarrett L. ....	28, 30
Jones, Philip W. ....	52
Joyce, John J. ....	134

## K

Kelly, Patrick ....	94
Kettering, Brett ....	82, 84
Khochfar, Sadegh ....	30
Kilcrease, David ....	102
Kumar, Rakesh ....	76

## L

Lamb, Kyle ....	84
Langdon, A. Bruce ....	106
Lawrence, Earl ....	92
Lecomte, Olivier ....	40
Leitner, Thomas ....	118
Lee, Christopher ....	100

Li, Hui.....	28, 30
Lipnikov, Konstantin.....	10, 20
Liu, Ke.....	70, 74
Livescu, Daniel.....	12, 14
Lowrie, Robert B. ....	4
Loxley, Peter N. ....	48
Lu, Chun-Yaung (Albert) .....	120

## M

Magee, Norman H. Jr.....	102
Malin, Alex .....	84
Maltrud, Mathew E.....	50, 52
Manzini, Gianmarco.....	44
Mara, Nathan A. ....	132
Mayeur, Jason R.....	132
McCabe, Rodney, J. ....	132
McClelland, Benjamin.....	68, 84
McClelland, Meghan .....	82
McMahon, Ben.....	116
Mernild, Sebastian H. ....	42
Mniszewski, Susan M. ....	86, 128
Mohd-Yusof, Jamaludin .....	88
Möller, Peter.....	108
Moulton, J. David.....	44
Mourad, Hashem M. ....	132
Muñoz, Dominic A. ....	6

Murillo, Michael S. ....	106
Mu, Fangping.....	112

## N

Nadiga, Balu T. ....	46, 48
Nesterov, Alexander I. ....	64
Niklasson, Anders.....	128
Nourgaliev, Robert R. ....	4
Nowak-Lovato, Kristy .....	112

## O

Olson-Reichhardt, Cynthia.....	142
Onunchi, Jose.....	122

## P

Painter, Scott.....	44
Pepper, Darrell W.....	8
Perez, Danny .....	120, 136
Peterson, Janet .....	58
Petersen, Mark R. ....	50, 52
Priedhorsky, Reid .....	86
Pritchard, Tyler .....	24
Pritchett-Sheats, Lori .....	4

## R

Rasmussen, Kim Ø. ....	112
Reichhardt, Charles.....	142

Reisner, Jon M. ....	16
Richards, David F. ....	106
Ribeiro, Ruy M. ....	114
Ringler, Todd ....	36, 52, 58
Rollin, Bertrand ....	16
Roming, Peter W.A. ....	24
Runde, Wolfgang H. ....	124

## S

Sambasivan, Shiv K. ....	18
Sanbonmatsu, Karissa ....	122
Sanville, Ed. ....	128
Shashkov, Mikhail ....	18
Sherrill, David ....	68
Sherrill, Manolo ....	102
Sims, Benjamin ....	94
Sintay, Stephen D. ....	132
Skar, Helena ....	118
Skurikhin, Mikhail ....	66
Smith, Pamela ....	68
Sridharan, Vilas ....	76
Straub, David N. ....	46
Svyatskiy, Daniil ....	20, 44
Surh, Michael P. ....	106
Swaminarayan, Sriram ....	88

## T

Thulasidasan, Sunil ....	90
Tian, Yong-Hui ....	124
Torres, Aaron ....	82
Torrez, Alfred ....	68
Tsifrinovich, Vladimir I. ....	98
Turner, Adrian K. ....	54

## U

Usheva, Anny ....	112
-------------------	-----

## V

Vander Weil, Scott ....	92
Vassilevki, Yuri ....	20
Volegov, Petr L. ....	98
Voter, Arthur F. ....	120, 136

## W

Wang, Jian ....	138, 140
Wang, Xiuling ....	8
Warren, Michael S. ....	32
Weaver, Brian ....	66
Wei, Tie ....	12, 14
Weijer, Wilbert ....	56
Wendelberger, Joanne ....	94
Whalen, Daniel J. ....	30

Williams, Sean J. ....	50
Womeldorff, Geoffrey A. ....	58

## Z

Zhang, Honglin ....	102
Zhang, Ruifeng ....	138
Zhang, Xiaodong ....	72
Zhang, Xuechen ....	70, 74
Zhou, Caizhi ....	140, 142
Zoglauer, Andreas ....	26
Zhu, Jian-Xin ....	134



## Organizational Abbreviations (for this publication only)

### **ADTSC**

Associate Directorate for Theory, Simulation, and Computation

### **ACS-P0**

Advanced Computing Solutions Program Office

### **Applied Computational Physics Division**

XCP-1 Lagrangian Codes

XCP-4 Methods and Algorithms

XCP-5 Materials and Physical Data

### **Applied Theoretical Design Division**

XTD-4 Improvised and Foreign Designs

### **Bioscience**

B-10 Biosecurity and Public Health

### **Chemistry Division**

C-IIAC Inorganic, Isotope, and Actinide Chemistry

### **Defense Systems and Analysis Division**

D-3 Systems Analysis and Surveillance

D-4 Energy and Infrastructure Analysis

### **Earth and Environmental Sciences Division**

EES-16 Computational Earth Sciences Group

### **Computer, Computational, and Statistical Sciences Division**

CCS-2 Computational Physics and Methods

CCS-3 Information Sciences

CCS-6 Statistical Sciences

CCS-7 Applied Computer Science

### **High Performance Computing Division**

HPC-1 Scientific Software Engineering

HPC-3 High Performance Computer Systems

HPC-5 System Integration

HPC-DO High Performance Computing Division Office

### **Intelligence, Analysis and Technology Division**

IAT-3 Threat Reduction Science and Engineering

### **LANL Institutes-National Security Education Center**

### **New Mexico Consortium**

### **Materials Physics and Applications Division**

MPA-CINT Center for Integrated Nanotechnologies

MPA-CMMS Condensed Matter and Magnet Science

### **Materials Science and Technology Division**

MST-6 Metallurgy

MST-8 Materials Science in Radiation and Dynamics Extremes

### **PADSTE**

Principal Associate Directorate for Science, Technology, and Engineering

### **Physics Division**

P-21 Applied Modern Physics

### **Theoretical Division**

CNLS Center for Nonlinear Studies

T-1 Physics and Chemistry of Materials

T-2 Nuclear and Particle Physics, Astrophysics and Cosmology

T-3 Fluid Dynamics and Solid Mechanics

T-4 Physics of Condensed Matter and Complex Systems

T-5 Applied Mathematics and Plasma Physics

T-6 Theoretical Biology and Biophysics

T-DO Theoretical Division Office

### **Technology Transfer Division**

TT-DO Technology Transfer

**DoD** Department of Defense

**DOE** Department of Energy

**NNSA** National Nuclear Security Administration

**ANL** Argonne National Laboratory

**INL** Idaho National Laboratory

**LLNL** Lawrence Livermore National Laboratory

**ORNL** Oak Ridge National Laboratory

**PNNL** Pacific Northwest National Laboratory

**SNL** Sandia National Laboratories

## Glossary Terms (for this publication only)

Term	Meaning
1D, 2D, 3D	One-, two-, or three-Dimensional
3-mer, 9-mer, 10-mer, 20-mer, k-mer	An n-gram of nucleic acid or amino acid sequences
AC	Alternating current
ACC	Antarctic Circumpolar Current
ALE	Arbitrary Lagrangian Eulerian
AoS	Array-of-Structures
AP	Abyssal Plain
APSP	All-Pairs Shortest-Path
APVM	Anticipated Potential Vorticity Method
AR5	regional division – Scandinavia
ARB	Accumulated Roll Bonding
ARPES	Angle-Resolved Photoemission Spectroscopy
ASC	Advanced Simulation and Computing Program
ASCEM	Advanced Simulation Capability for Environmental Management
ATS	Arctic Terrestrial Simulator
AU	Astronomical Unit, the distance between the Earth and the Sun
bcc	Body-Centered Cubic
BH	Black Hole
BHR	Besnard Harlow Rauenzahn
BIOBASE	Biological Database
BMSD	Bulk-Mediated Surface Diffusion
C	Charge Conjugation Capacitance
C-2	Clements
CaCO <sub>3</sub>	Calcium Carbonate
CASL	Consortium for Advanced Simulation of Light Water Reactors
CDC	Centers for Disease Control
CESM	Climate Earth System Model
CMB	Cosmic Microwave Background
Co	Cobalt

CoMD	Co-Designed Molecular Dynamics
COP	Characteristic Opening Profile
COSIM	Climate, Ocean, and Sea Ice Modeling
CP	Charge Conjugation–Mirror Reflection
CPU	Central Processing Unit
CRADA	Cooperative Research and Development Agreement
CRF	Circulating Recombinant Form
Cu	Copper
DCC	Double chipkill correct memories
DES	Dark Energy Survey
DFT	Density Functional Theory
DIMM	Dual In-Line Memory Module
DMFT	Dynamical Mean-Field Theory
DMP	Discrete Maximum Principle
DNA	Deoxyribonucleic Acid
DNS	Direct Numerical Simulation
DOS	Density of States
DP	Detection Probabilities
DRAM	Dynamic Random-Access Memory
DSL	Domain-Specific Language
EAM	Embedded Atom Method
EBS	Electron backscatter defraction
ECC	Error Correcting Codes
ECW	Electrochemical Window
EDC	Energy Distribution Curve
EMSA	Electrophoretic Mobility Shift Assay
EOS	Equation-of-State
ET	Electron Transfer
ExMatEx	Exascale Co-design Center for Materials in Extreme Environments
fcc	Face-Centered Cubic
FD	Finite Difference
FE	Finite Element

Fe	Iron
FEM	Finite Element Method
Fis	Nucleoid-associated Protein in Escherichia coli
FMO	Fenna-Matthews-Oslov
FRD	Fast Radiation Damping
FS	Ackland Finnis-Sinclair
FV	Finite Volume
FY	Fiscal Year
FZ	Fracture Zone
Ga	Gallium
GB	Gigabyte
Genbank	Sequence Database of Nucleotide Sequences and their Protein Translations
GIC	Glacier and ice caps
GM	Gent-McWilliams Closure
GM	Gent-McWilliams Isopycnal Transport and Mixing Scheme
GPU	Graphics Processing Unit
GRACE	Gravity Recovery and Climate Experiment
GSF	Generalized Stacking Fault
GTRF	Grid-to-rod fretting
H	Hydrogen
H1N1	Influenza A
H2	Hydrogen
H2N2	Influenza A Subtype
H5N1	Avian Influenza
Hae-Rsa	Restriction Fragment
HDFS	Hadoop Distributed Filesystem
He	Helium
HIV	Human Immunodeficiency Virus
HIV-1	Human Immunodeficiency Virus Type 1
HMX	Energetic Component of PBX
HNC	Hypernetted Chain Approximation
hp-adaptive	Method employing hierarchical basis functions
HPC	High Performance Computing

I/O	Input/Output
ICF	Inertial Confinement Fusion
IEEE	Institute of Electrical and Electronics Engineers
IgG	Immunoglobulin G
ILES	Implicit Large Eddy Simulations
ILI	Influenza-like Illness
In	Indium
I/OFSL	Input/Output Forwarding Scalability Layer
IP	Internet Protocol
IPCC	Intergovernmental Panel on Climate Change
IT	Information Technology
JMP	DoD/DOE Joint Munitions Program
JWST	James Webb Space Telescope
K	Potassium
KPM	Kernel Polynomial Method
KPSS	Konstandin, Prokopec, Schmidt, and Seco
KS	Kurdjumov Sachs
L	Inductance
LBN	Logical Block Number
LCRS	Inductance, Capacitance, Resistance, Switch
LDA	Local density appropriation
LES	Large Eddy Simulation
LHC	Light-Harvesting Complex
LSST	Large Synoptic Survey Telescope
LSA	Linear Stability Analysis
LST	Linear Stability Theory
LTE	Local Thermodynamic Equilibrium
M&HS	Combined Facemasks and Hand Sanitizers
M&SD	Combined Facemasks and Social Distancing
Mach	Speed Measurement
MAID	Massive Array of Idle Disk
MB	Material Balance
MD	Molecular Dynamics
MD	Molecular Dynamics
MDC	Momentum Distribution Curve



MEP	Mesoscale Eddy Parametrization
MFD	Mimetic Finite Difference
MFE	Mixed Finite Element
Mg	Magnesium
MIC	Many Integrated Core
MP	Multi-point
MPAS	Model Prediction Across Scales
MPC	Multi-Process Coordinators
MPFA	Multi-Point Flux Approximation
MPI/IO	Message passing interface/input output
mRNA	Messenger RNA
MSD	Mean Squared Displacement
n-gram	Contiguous Sequence of n Items from a Given Sequence of Text or Speech
N5N1	Avian Influenza
N95	respirators
NAIM	Nucleotide Analog Interference Mapping
NAS	National Academy Study
Nb	Niobium
NCAR	National Center for Atmospheric Research
NCBI	National Center for Biotechnology Information
Ne	Neon
Ni	Nickel
NICS	National Institute for Computational Sciences
NMA	Nuclear Materials Accounting
NMR	Nuclear Magnetic Resonance
NS oligo	Oligonucleotide Competitor
NuSTAR	Nuclear Spectroscopy Telescope Array
ONIOM	Our own N-layered Integrated molecular Orbital and Molecular mechanics
OP	Opacity Project
OPAL	Opacity Database at LLNL
OPLIB	Opacity Database at LANL
OS	Operating System CAD
P	Mirror Reflection

P-G	Petrov-Galerkin unwinding technique
PB	Poisson-Boltzmann
PBM	Protein Binding Microarrays
PBX	Plastic-Bonded Explosive
PC	Photosynthetic Complex Polycarbonate
PCM	Polarizable Continuum Model
PCR	Polymerase Chain Reaction
PCS	Predictor-Corrector Scheme
Pd	Palladium
PE	Variance of the Potential Vorticity
PK	Process Kernels
PLFS	Parallel Log File System
PLG	Human Promoter, DNA
PM	Process Monitoring
Pop III	Primordial Star
PSI	Photosystem I
PSII	Photosystem II
Pt	Platinum
Pu	Plutonium
PV	Potential vorticity
PWM	Position Weight Matrix
QM/MM	Quantum mechanics/molecular mechanics
QMC	Quantum Monte Carlo
QoS	Quality of service
R	Resistance
RAM	Random Access Memory
RANS	Reynolds Averaged Navier-Stokes
RAVE	Reconfigurable Advanced Visualization Environment
RC	Reaction Center
ReALE	Reconnection-Arbitrary-Lagrangian-Eulerian
REP	Perturbation Reynolds number
rf	Radio Frequency
Rh	Rhodium
RNA	Ribonucleic Acid

RNase	Ribonuclease
RPA	Random Phase Approximation
RPM	Revolutions Per Minute
rRNA	Ribosomal RNA
Rsa-Mse	Restriction Fragment
Rsun	Radius of the sun
RSV	Respiratory Syncytial Virus
RTI	Rayleigh-Taylor Instability
RTIL	Room-Temperature Ionic Liquid
S	Switch
SAM	S-adenosylmethionine
SARS	Severe Acute Respiratory Syndrome
SCDCD	Probability of uncorrectable error
SCVT	Spherical Centroidal Voronoi Tessellation
SECDEC	Single Error Correction, Double Error Detection
SGH	Staggered Grid Hydrodynamics
SGS	Subgrid scale
SHAPE	Selective 2'-Hydroxyl Acylation by Primer Extension
Si	Silicon
SIMD	Single Instruction Multiple Data
SIR	Susceptible, infectious, recovered
SKA	Square Kilometer Array
SLE	Sea Level Equivalent
SMBH	Supermassive Black Hole
SMS	Supermassive Star
SN	Supernova
SNe	Supernovae
SoA	Structure-of-Arrays
SPD	Severe Plastic Deformation
SPT	South Pole Telescope
ssChIP	Site-Specific Chromatin Immunoprecipitation
SSD	Solid State Drives
SSH	Secure Shell
SSH	Sea Surface Height

SSSP	Single-Source Shortest-Path
STEM	Science, Technology, Engineering, and Mathematics
TDC	Thermodynamic Cycle
TF	Transcription Factor
Ti	Titanium
TSG	Telescoping Subgraph
TSS	Transcription Start Site
TTS	Temporary Triangular Subzoning
U	Uranium
UE	Unrecoverable error
UQ	Uncertainty Quantification
US	United States
UV	Ultraviolet
VD	Variable Density
VLSI	Very Large Scale Integration
YeTI	Yellow-to-Turquoise Integration
YY1	Human Transcription Factor
Zn	Zinc







